Computing optical properties of large systems

Tim Joachim Zuehlsdorff

Department of Physics

Imperial College London

September 2014

A thesis submitted for the degree of Doctor of Philosophy of Imperial College London and Diploma of the Imperial College

Acknowledgements

The research outlined in this dissertation was carried out through the EPSRC Centre for Doctoral Training on Theory and Simulation of Materials at Imperial College London.

I would like to thank my supervisors Peter Haynes, James Spencer, Nicholas Harrison and Jason Riley for their constant support, helpful advice and stimulating discussions over the last three years.

I am indebted to Nick Hine for furthering my understanding of the inner workings of the `ONETEP` code as well as the PAW method.

Niccolo Corsini and Daniel Cole provided some of the atomistic structures of the systems studied in this work.

Further thanks are due to members of my cohort at the TSM CDT for creating a fun and stimulating learning environment. A special thanks goes out to Tom Poole, David Trevelyan, Nina Kearsey, Fabian Renn and Benat Gurrutxaga-Lerma for providing much needed distractions in the form of a game of catch in the corridor of the Whiteley Suite.

Thanks to everyone who joined me for an after work pint on the the occasional more frustrating days filled with compiler errors and segmentation faults, especially Red, Rob, Stefan, Ruben and Yannick.

Furthermore, I would like to thank every member of the Imperial Eagles Handball Club. There is no one I would rather spend an early Saturday morning trip to Barking or an overnight twelve hour minibus journey to Lyon with than you people.

Finally, I thank my family, both for their love and constant support and for giving me the opportunity to embark on that journey I started nine years ago by coming to the UK.

Declaration of Originality

I declare that the work contained within this thesis is my own.
It has not been submitted for any award elsewhere.
Figures 6.3, 6.5 and 6.8, as well as table 6.2 have already appeared in an earlier publication [99] and are reproduced here with permission of AIP Publishing.

Declaration of Copyright

Abstract

In recent years, time-dependent density-functional theory (TDDFT) has been the method of choice for calculating optical excitations in medium sized to large systems, due to its good balance between computational cost and achievable accuracy. In this thesis, TDDFT is reformulated to fit the framework of the linear-scaling density-functional theory (DFT) code `ONETEP`. The implementation relies on representing the optical response of the system using two sets of localised, atom centered, *in situ* optimised orbitals in order to ideally describe both the electron and the hole wavefunctions of the excitation. This dual representation approach requires only a minimal number of localised functions, leading to a very efficient algorithm. It is demonstrated that the method has the capability of computing low energy excitations of systems containing thousands of atoms in a computational effort that scales linearly with system size.

The localised representation of the response to a perturbation allows for the selective convergence of excitations localised in certain regions of a larger system. The excitations of the whole system can then be obtained by treating the coupling between different subsystems perturbatively. It is shown that in the limit of weakly coupled excitons, the results obtained with the coupled subsystem approach agree with a full treatment of the entire system, with a large reduction in computational cost.

The strengths of the methodology developed in this work are demonstrated on a number of realistic test systems, such as doped *p*-terphenyl molecular crystals and the exciton coupling in the Fenna-Matthews-Olson complex of bacteriochlorophyll. It is shown that the coupled subsystem TDDFT approach allows for the treatment of system sizes inaccessible by previous methods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."

Paul Dirac [1]

1.1 Practical methods

When the theory of quantum mechanics was developed in the early 20th century, it caused the most groundbreaking revolution in scientific thinking since the publication of Newton's *Principia* over 200 years prior. It was quickly recognised that the theory did in principle allow for the calculation of all properties of matter from the atomic scale all the way up to the macroscopic scale. Or, in Dirac's perhaps slightly optimistic words, quantum mechanics provided a mathematical theory for "a large part of physics and the whole of chemistry". However, it soon became equally apparent that the equations describing the fundamental behaviour of matter were far too complicated to be soluble for more than a handful of the most trivial cases.

For this reason, ever since the development of the basic theory of quantum mechanics, a significant research effort has been directed into developing "practical methods" that would allow for calculating approximate solutions to complex systems. The earliest approximations, like the WKBJ method (named after Wentzel, Kramers, Brillouin and Jeffreys, who each independently developed it) [2–5], focused on finding so-called "semi-classical solutions" to the Schrödinger equation, the main equation governing quantum mechanics. It has been used extensively in finding the behaviour of quantum particles in simplified models of tunneling processes, most notably to derive the rate of nuclear

fusion [6].

The introduction of computers helped along a further development in quantum mechanics, namely the move away from studying simplified model systems towards finding approximate numerical solutions to real systems like molecules and solids. To unlock the predictive power of quantum mechanics, it becomes necessary to make use of methods that are truly *ab initio*, in that they do not make use of any *a priori* assumptions about behaviour of the system that is studied. However, the complexity of the quantum mechanical equations that Dirac already recognised in 1926 means that a direct numerical solution requires a computational effort that grows exponentially with the size of the system that is to be solved, rendering any such direct approach impractical for anything but the most simple problems. Thus in practice, further well-controlled approximations have to be made in order to reduce the scaling from an exponential to a polynomial one.

The most successful in this new generation of "practical methods" are based on density-functional theory (DFT) [7], which after making a number of simplifying approximations to the quantum mechanical effects of exchange and correlation of electrons, yields any ground state property of a system with a computational scaling that is just the cube of the system size. Since its development in 1964 it has shown real predictive power in the study of molecules and solids and has become the standard tool in areas as diverse as chemistry, biophysics and materials science. The great impact of DFT on a wide range of scientific disciplines was recognized in 1998 by awarding the Nobel Prize of Chemistry to Walter Kohn and John Pople.

1.2 Spanning lengths

While the standard formulation of DFT has been remarkably successful in the last few decades, measured in both the annual citations of the papers describing the original method and the number of different scientific areas it is now routinely used in, most of its applications are focused on two limiting cases. On one end of the scale are small, isolated molecules, while the other end of the scale contains applications to infinite crystals. The reason why most applications are limited to these two regimes can be seen in the $\mathcal{O}(N^3)$ scaling with system size that is inherent in the method, meaning that a doubling of system size leads to an 8-fold increase in computational effort. Both the case of a small, isolated molecule and an infinite (defect-free) crystal can be computed by considering only a few atoms[1] and are easily treated by standard DFT with moderate computational effort.

In recent years there has been an increased interest in systems that lie between these

[1] In the case of infinite systems, this is achieved by making use of periodic boundary conditions, such that only a unit cell of the crystal has to be treated explicitly and the translational crystal symmetry is exploited.

two limiting cases and that are prohibitively large for the $\mathcal{O}(N^3)$ scaling of DFT. Examples of these systems include large biomolecules, nanocrystals and infinite crystals containing defects. The feature all of these systems have in common is that an appropriate treatment of them requires calculations containing thousands of atoms, rendering them impractical using conventional DFT even on modern supercomputers.

Over the last 20 years, linear-scaling DFT methods have been designed specifically to address the class of systems mentioned above and are now routinely used to treat systems containing thousands of atoms. This new generation of methods has been used with great success to treat problems that were completely inaccessible even a few years ago.

1.3 Light interactions

While the development of linear-scaling techniques has opened up a wide field of new applications for DFT (see, for example [8–11]), it also comes with a new set of challenges. Since calculations of ground state properties of systems containing thousands of atoms are now possible, the problem shifts to extracting information from these calculations that can be compared directly to experimental measurements.

Consider for example a crystal with two interacting defects in it. In order to find the lowest energy arrangement of the defects in the entire structure using linear-scaling DFT one has to generate all unique configurations of the two defects in the lattice and perform a full DFT calculation on every one of these structures to compare the total energies. While each of the individual DFT calculations does scale linearly with the number of lattice sites within a finite supercell, the number of nonequivalent defect configurations grows as $\mathcal{O}(N)$ with lattice sites, leading to an overall scaling of $\mathcal{O}(N^2)$. Thus in this example the scaling problem does not originate from the DFT calculation itself but from the configurational complexity that increases once the system size is increased. To make matters worse, the above example of two interacting defects is likely to produce a large array of potential structures, all separated by very small energy differences, any of which are likely to occur in the real material whose properties need to be predicted. Due to this configurational complexity in large systems, it is often found that predicting ground state properties of the single lowest energy configuration is of little use in making predictions of the behaviour of the entire system.

A way of sidestepping the configurational complexity problem leading to unfavourable scaling when linear-scaling DFT is applied to a large system is to attempt to predict properties that are measured directly in experiments. One of the properties of interest for many practical applications is the interaction of the system with light, which can be obtained from the time-dependent extension of the DFT method (TDDFT). Accurate

predictions of absorption spectra of semiconductor nanocrystals are a key ingredient to developing new generations of efficient solar cells, while predictions of spectra of large photoactive biomolecules is of great interest in the area of biophysics.

Optical absorption spectra of nanostructures and large photoactive biomolecules that are measured in experiments are taken on timescales that are several orders of magnitude larger than the timescales of atomic vibrations around their point of equilibrium. Thus the optical spectra produced in experiments can be seen as the result of an averaging over many low energy configurational structures of the system of interest and are often well reproduced by theoretical calculations on a single, average low energy structure of that system[2]. Thus in order to make meaningful predictions on the measurable quantity of light absorption, the configurational complexity problem that troubles many potential applications of linear scaling techniques can be, to a large extent, ignored, making theoretical spectroscopy an ideal area of impact for linear-scaling methods.

The purpose of this dissertation is to extend the linear-scaling techniques that proved so successful in standard DFT over recent years to the calculation of optical spectra using TDDFT. The aim is to develop methods capable of calculating the low energy optical spectrum of systems containing thousands of atoms in an effort scaling linearly with system size. This will open up new potential areas of research in the field of theoretical spectroscopy and connect directly to experimental measurements to enable a more effective collaboration between the experimental and the theoretical community.

1.4 Dissertation outline

This dissertation is organised in the following way.

Chapter 2 contains a brief discussion of some of the prerequisite knowledge for the later chapters. Some basic concepts of quantum mechanics are introduced, as well as the the main type of system that is solved for in later chapters and some conventions in the scientific notation that is used throughout this work.

In chapter 3 the theoretical foundations of density-functional theory are discussed. The focus of the chapter is on introducing the main approximations and techniques used in practical DFT calculations, both for periodic infinite systems and localised finite systems. Finally, the main concepts of density-matrix DFT are introduced and it is demonstrated

[2]It should be noted that such an average, low energy structure cannot always be readily found in large biological systems where atomic positions are derived from X-ray diffraction experiments. Furthermore, while atoms in nanostructured crystals often undergo simple oscillations around a point of equilibrium, this is not necessarily the case in biological pigment-protein complexes, where the motion of the protein happening on a much longer timescale can become important. In these systems, it is often necessary to calculate spectra of several snapshots taken from a molecular dynamics simulation in order to achieve a good representation of the relevant phase space.

how this method can yield algorithms scaling linearly with system sizes.

In chapter 4, the focus is shifted from ground state properties to excited state properties. The formal justification for time-dependent density-functional theory is introduced before moving on to practical implementations. Both the advantages and disadvantages of the two main flavours of TDDFT, time-evolution TDDFT and linear-response TDDFT, are discussed in the light of the objective of this dissertation. Finally, the TDDFT method is compared and contrasted to a different approach to the excited state problem, namely many-body perturbation theory, where the focus is on the so-called GW-approximation and the Bethe-Salpeter equation.

Chapter 5 gives a brief overview over some specific features of the linear-scaling DFT code `ONETEP` that are relied upon in later chapters. Here, the focus is on properties needed for the derivation of the linear-scaling TDDFT method in the next chapter, as well as some specific functionality going beyond standard DFT that is used in later benchmark calculations.

Chapter 6 contains the derivation of the linear-scaling TDDFT method developed in this dissertation. An algorithm is introduced that is capable of obtaining low-energy excitations of large systems in linear-scaling effort and its convergence properties are discussed. A number of benchmark calculations are performed and an excellent agreement with conventional TDDFT codes is found. Furthermore, it is demonstrated on a test system that the algorithm indeed scales linearly with system size.

In chapter 7 a number of appropriate changes to the linear-scaling TDDFT algorithm are derived to make it compatible with the projector augmented-wave (PAW) formalism used in ground-state calculations of transition metal elements in `ONETEP`. The correct behaviour of the code is demonstrated by comparing to all-electron TDDFT results and a proof-of-priniple calculation of the low-energy optical spectrum of a Cadmium-Sulphide (CdS) nanocrystal under pressure is performed.

In chapter 8 it is demonstrated that placing constraints on the sparsity of the response density matrix allows for the calculation of targeted excitations localised to certain subsystems of a larger system. It is furthermore shown that the interaction between subsystems can be introduced perturbatively, yielding the excitation spectrum of the entire system to a good degree of accuracy in the limit of weakly coupled excitons.

Finally, in chapter 9, some practical uses of the method developed in this work are shown, where two physically motivated, large scale applications are considered. First, the exciton coupling between chromophores in the Fenna-Matthews-Olson light-harvesting complex is calculated, demonstrating both the importance of large-scale calculations to correctly capture environmental effects and the power of the subsystem TDDFT technique to speed up calculations. The other large scale system treated is a p-terphenyl organic

crystal doped with different acene derivatives. It is demonstrated that linear scaling TDDFT can be used as a part of a prescreening process to select potential dopants for applications in room-temperature masers.

The dissertation is concluded with chapter 10, summarising the main findings. Finally, some potentially interesting further research is pointed out, with the capability to expand on the scope of the work discussed here.

Chapter 2

Theoretical background: Prerequisites

In this chapter, some of the prerequisite background knowledge that is needed for the later chapters is discussed. The chapter gives a very brief overview over some important aspects of quantum mechanics and introduces the most important choices of notation that are used throughout the rest of the work.

2.1 The wavefunction

The wavefunction Ψ completely describes the quantum state of a system of N particles and contains all information about the system. While the wavefunction does not have a unique representation for an arbitrary quantum system, a typical representation for an N-particle system is the position representation, in which Ψ can be written as a complex $3N+1$ dimensional function[3] $\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \cdots \mathbf{r}_N, t)$ of the positions of all N particles and some time t. The square modulus of the wavefunction $\Psi^*\Psi$ can be interpreted as a probability distribution function of the N-particle system. In order for this interpretation to be valid, the wavefunction is required to follow normalisation conditions, which, for an N-particle system in position representation, can be suitably written as:

$$\int \mathrm{d}^3 r_1 \mathrm{d}^3 r_2 \cdots \mathrm{d}^3 r_N \, \Psi^*(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \cdots \mathbf{r}_N, t)\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \cdots \mathbf{r}_N, t) = 1 \tag{2.1}$$

For most of this work the Dirac notation will be used, where the wavefunction is denoted as $|\Psi\rangle$, a state vector in the Hilbert space. The advantage of the Dirac notation can be seen in that it is representation independent and thus provides a generalised way of introducing quantum mechanical concepts. In Dirac notation, the normalisation constraint of equation (2.1) can be simply written as $\langle\Psi|\Psi\rangle = 1$. The existence of a wavefunction with the above properties for any arbitrary system forms the first postulate of quantum mechanics [12].

[3] This notation implies that an intrinsic property of quantum mechanical particles called *spin* is ignored for the time being. It is formally introduced later in this chapter.

2.2 Operators and basis functions

The second postulate of quantum mechanics states that every classical (ie. physical) observable has a corresponding linear operator [12]. Considering an operator $\hat{A}$ with associated observables $\{a_n\}$ one can write

$$\hat{A}|\psi_n\rangle = a_n|\psi_n\rangle \tag{2.2}$$

where ψ_n is said to be an eigenfunction of operator $\hat{A}$ with an associated observable eigenvalue a_n. All operators of observables are required to be linear, from which it follows that if there are two eigenstates $|\psi_1\rangle$ and $|\psi_2\rangle$ of operator $\hat{A}$ with the *same* eigenvalue a, then every linear combination of $|\psi_1\rangle$ and $|\psi_2\rangle$ is also an eigenstate of $\hat{A}$ with eigenvalue a, since

$$\hat{A}(c_1|\psi_1\rangle + c_2|\psi_2\rangle) = c_1\hat{A}|\psi_1\rangle + c_2\hat{A}|\psi_2\rangle = a(c_1|\psi_1\rangle + c_2|\psi_2\rangle). \tag{2.3}$$

Furthermore, all operators associated with classical observables are required to be Hermitian or self-adjoint, ie. $\hat{A}^\dagger = \hat{A}$. This equality is required such that

$$\langle\psi_i|\hat{A}|\psi_i\rangle^* = \langle\psi_i|\hat{A}^\dagger|\psi_i\rangle = \langle\psi_i|\hat{A}|\psi_i\rangle \tag{2.4}$$

from which it follows immediately that the eigenvalues a_n of an Hermitian operator are real, as required for observable properties of a system. The Hermitian property has one further consequence for the eigenstates of quantum mechanical operators. Consider two states $|\psi_i\rangle$ and $|\psi_j\rangle$ such that $\hat{A}|\psi_j\rangle = a_j|\psi_j\rangle$ and $\hat{A}|\psi_i\rangle = a_i|\psi_i\rangle$. Operating with $\langle\psi_i|$ and $\langle\psi_j|$ on those two expressions respectively one obtains

$$\begin{aligned} & \langle\psi_i|\hat{A}|\psi_j\rangle = a_j\langle\psi_i|\psi_j\rangle \\ & \langle\psi_j|\hat{A}|\psi_i\rangle = a_i\langle\psi_j|\psi_i\rangle \\ \Rightarrow \quad & \langle\psi_i|\hat{A}|\psi_j\rangle - \langle\psi_j|\hat{A}|\psi_i\rangle = (a_j - a_i^*)\langle\psi_i|\psi_j\rangle. \end{aligned} \tag{2.5}$$

However, due to the Hermitian properties of $\hat{A}$, the left hand side of the equation must vanish, such that one obtains

$$(a_j - a_i^*)\langle\psi_i|\psi_j\rangle = 0. \tag{2.6}$$

If a_i and a_j are non-degenerate, then it is required that $\langle\psi_i|\psi_j\rangle = 0$ and thus the two eigenstates are orthogonal to each other. Thus eigenstates of Hermitian operators with different eigenvalues are necessarily orthogonal to each other.

Due to the orthogonality property the eigenfunctions of an Hermitian operator form a complete basis and can be used to expand an arbitrary quantum state $|\Psi\rangle$ of the system.

Written in this basis the wavefunction becomes

$$\begin{aligned} |\Psi\rangle &= \sum_{n=0}^{\infty} c_n |\psi_n\rangle \\ c_n &= \langle \Psi | \psi_n \rangle^* \end{aligned} \tag{2.7}$$

Any measurement of the physical quantity associated with operator $\hat{A}$ yields the quantity a_n with a probability of $|c_n|^2$. Thus if all the eigenstates and eigenvalues of a Hermitian operator are known, the probability of a given measurement outcome on an arbitrary quantum state $|\Psi\rangle$ can be predicted. If $|\Psi\rangle$ can be represented by a single $|\psi_n\rangle$, it follows that $|\Psi\rangle$ is an eigenstate of $\hat{A}$ and any measurement will yield a_n. Furthermore, for an arbitrary normalised state $|\Psi\rangle$, the quantity $\langle \Psi | \hat{A} | \Psi \rangle$ is the expectation value of a measurement of the observable associated with $\hat{A}$.

Since a measurement of the observable associated with $\hat{A}$ yields a single well-defined outcome a_n if and only if $|\Psi\rangle = |\psi_n\rangle$, it follows that the measurement of a second observable associated with $\hat{B}$ simultaneously can only yield a single well-defined outcome if $|\psi_n\rangle$ is also an eigenstate of $\hat{B}$. Writing $|\psi_n\rangle$ as $|\psi_{n,m}\rangle$ to signify that it has associated eigenstates a_n and b_m, one can write

$$\hat{A}\hat{B}|\psi_{n,m}\rangle = \hat{B}\hat{A}|\psi_{n,m}\rangle = a_n b_m |\psi_{n,m}\rangle \tag{2.8}$$

from which it follows that

$$\left(\hat{A}\hat{B} - \hat{B}\hat{A}\right)|\psi_{n,m}\rangle = \left[\hat{A}, \hat{B}\right]|\psi_{n,m}\rangle = 0. \tag{2.9}$$

The quantity $\left[\hat{A}, \hat{B}\right]$ is known as the commutator. Thus, if the commutator of two operators vanishes, the two operators share a common set of eigenstates and the two observables associated with $\hat{A}$ and $\hat{B}$ can be measured simultaneously with arbitrary precision.

2.3 The Schrödinger equation

The quantum mechanical operator related to the classical observable of the total energy of a system is the Hamiltonian, commonly denoted as $\hat{H}$. It is most commonly written as $\hat{H} = \hat{T} + \hat{V}$, ie. as the sum of the kinetic energy operator and the potential energy operator. In quantum mechanics, the Hamiltonian is of fundamental importance since it generates the time evolution of a quantum state $|\Psi\rangle$ via the time-dependent Schrödinger

equation, which can be written as:

$$\hat{H}|\Psi\rangle = \mathrm{i}\hbar\frac{\partial}{\partial t}|\Psi\rangle \tag{2.10}$$

where $\hbar$ is Planck's constant divided by 2π. Solving the Schrödinger equation for the Hamiltonian of a given system yields the full time evolution of any eigenstate of said Hamiltonian. Here, a special case of the above equation is considered, that occurs when $\hat{H}$ is time-independent. If $\hat{H}$ has no explicit dependence on time one can use the following Ansatz for the quantum state (working explicitly in a position representation for the time being): $\Psi(\mathbf{r}_1 \cdots \mathbf{r}_n, t) = \psi(\mathbf{r}_1, \cdots \mathbf{r}_n)\tau(t)$. Using the Ansatz, the spatial coordinates are separated from the time coordinate. Since $\hat{H}$ is time-independent, it only operates on $\psi(\mathbf{r}_1, \cdots \mathbf{r}_n)$, while the partial derivative on the right hand side of equation (2.10) only operates on $\tau(t)$. It follows that

$$\psi^{-1}(\mathbf{r}_1, \cdots \mathbf{r}_n)\hat{H}\psi(\mathbf{r}_1, \cdots \mathbf{r}_n) = \mathrm{i}\hbar\tau^{-1}(t)\frac{\partial\tau(t)}{\partial t} \tag{2.11}$$

and from this it is found that

$$\hat{H}\psi(\mathbf{r}_1, \cdots \mathbf{r}_n) = E\psi(\mathbf{r}_1, \cdots \mathbf{r}_n) \tag{2.12}$$

$$\Psi(\mathbf{r}_1, \cdots \mathbf{r}_n, t) = \psi(\mathbf{r}_1, \cdots \mathbf{r}_n)e^{-\mathrm{i}Et}. \tag{2.13}$$

Equation (2.12) is known as the time-independent Schrödinger equation. Switching back to Dirac notation, the above expressions state that for a time-independent Hamiltonian, in order to find the time-evolution of any quantum state, it is sufficient to solve $\hat{H}|\psi_n\rangle = E_n|\psi_n\rangle$ for $\{E_n\}$ and $\{|\psi_n\rangle\}$, the set of all eigenstates and eigenvalues of the Hamiltonian.

2.4 Symmetry and spin

One major postulate of quantum mechanics is that quantum particles described by the wavefunction are indistinguishable. This means that for a two-particle wavefunction, the states $\Psi(\mathbf{r}_1, \mathbf{r}_2)$ and $\Psi(\mathbf{r}_2, \mathbf{r}_1)$ must describe the same physical system. Therefore, the wavefunctions under particle exchange can at most differ by some complex constant c. Defining the particle interchange operator $\hat{P}_{12}$ such that

$$\hat{P}_{12}|\Psi(\mathbf{r}_1, \mathbf{r}_2)\rangle = |\Psi(\mathbf{r}_2, \mathbf{r}_1)\rangle \tag{2.14}$$

it follows that

$$\hat{P}_{12}\hat{P}_{12}|\Psi(\mathbf{r}_1,\mathbf{r}_2)\rangle = \hat{I}|\Psi(\mathbf{r}_1,\mathbf{r}_2)\rangle = |\Psi(\mathbf{r}_1,\mathbf{r}_2)\rangle \tag{2.15}$$

where $\hat{I}$ is the identity operator. The above considerations require that the eigenvalues of $\hat{P}_{12}$ are ± 1 and that one can therefore write

$$\Psi(\mathbf{r}_1,\mathbf{r}_2) = \pm\Psi(\mathbf{r}_2,\mathbf{r}_1). \tag{2.16}$$

The above result can be generalised to any number of particles such that the total wavefunction is either symmetric or antisymmetric. Whether it is symmetric or antisymmetric follows from an intrinsic angular momentum of quantum mechanical particles called spin, that does not have a classical equivalent and is a consequence of a relativistic extension of the Schrödinger equation [13]. Quantum mechanical particles come in two basic flavours: bosons with an intrinsic spin equal to an integer number are described by a symmetric wavefunction, while fermions with half-integer spin are described by antisymmetric wavefunctions.

To correctly account for spin, it is neccessary to explicitly include it in the wavefunction. One thus introduces the set of collective variables $\{\mathbf{x}_i\} = \{\mathbf{r}_i, \sigma_i\}$ containing a position vector and spin index for each particle. In this work the only quantum mechanical particles of interest are electrons, which have a spin of 1/2. This leads to their wavefunctions being antisymmetric, ie.

$$\Psi(\cdots,\mathbf{x}_i,\cdots,\mathbf{x}_j,\cdots) = -\Psi(\cdots,\mathbf{x}_j,\cdots,\mathbf{x}_i,\cdots) \tag{2.17}$$

An important consequence of the antisymmetry condition of the wavefunction is that two indistinguishable fermions cannot occupy the same quantum state, a property known as the Pauli exclusion principle.

Even without considering the full relativistic Dirac equation, the effects of the spin dependence of the many-electron wavefunction can often be introduced into the Hamiltonian in an approximate form, or treated perturbatively. For the purpose of this work however no explicit spin dependence of the Hamiltonian is considered. Thus the main consequence of the electron spin in this work originates from the symmetry restrictions of the wavefunction and the Pauli exclusion principle only.

2.5 The variational principle

Consider a Hamiltonian $\hat{H}$ with associated energies and eigenstates $\{E_n\}$ and $\{|\psi_n\rangle\}$. Now consider the system to be in an arbitrary, normalised state $|\Psi\rangle$. Then the variational principle states that the expectation value of an energy measurement of the system in

state $|\Psi\rangle$ has to follow

$$\langle\Psi|\hat{H}|\Psi\rangle \geq E_0 \tag{2.18}$$

with the equality only reached if $|\Psi\rangle = |\psi_0\rangle$.

This can be seen by expanding $|\Psi\rangle = \sum_{n=0}^{\infty} c_n\psi_n$ and using the orthogonality of the eigenstates of $\hat{H}$:

$$\begin{aligned} E[\Psi] &= \langle\Psi|\hat{H}|\Psi\rangle \\ &= \sum_{i,j=0}^{\infty} c_i^* c_j E_j \langle\psi_i|\psi_j\rangle \\ &= \sum_{i=0}^{\infty} |c_i|^2 E_i \end{aligned} \tag{2.19}$$

Finally, subtracting E_0 from both sides one obtains

$$E[\Psi] - E_0 = \sum_{i=0}^{\infty} |c_i|^2 (E_i - E_0). \tag{2.20}$$

Since every term on the right hand side of the equation is greater than or equal to zero, $E[\Psi] - E_0 \geq 0$ and the equality only holds if $c_1, c_2 \cdots c_\infty = 0$ (provided E_0 is non-degenerate).

The variational method is often used to compute the ground state of a quantum mechanical system. The principle allows for the calculation of the ground state energy of a system by starting from a trial wavefunction $|\Psi\rangle$ and then minimising the functional $E[\Psi]$ with respect to Ψ. If a minimum is reached, by the variational principle $\Psi = \psi_0$ and $E[\Psi] = E_0$, provided the ground state is non-degenerate.

2.6 The molecular Schrödinger equation and Hartree atomic units

At this point, the exact form of the Hamiltonian for a system of interacting atoms is introduced. For reasons of simplicity, the position representation is used throughout, the nucleic positions are denoted as $\mathbf{R} \equiv \{\mathbf{R}_A\}$ and the electron positions as $\mathbf{r} \equiv \{\mathbf{r}_i\}$. Ignoring the electron spin for the moment, the time-independent part of the wavefunction of the system of nuclei and electrons can be written as $\Psi = \Psi(\mathbf{R}, \mathbf{r})$. The time-independent

molecular Hamiltonian of the system can then be written as:

$$\begin{aligned}\hat{H} &= -\sum_i \frac{1}{2m_e}\nabla_i^2 - \sum_A \frac{1}{2M_A}\nabla_A^2 - \sum_{i,A} \frac{Z_A e^2}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i>j} \frac{e^2}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}_j|} \\ &\quad + \sum_{A>B} \frac{Z_A Z_B e^2}{4\pi\epsilon_0|\mathbf{R}_A - \mathbf{R}_B|}\end{aligned} \tag{2.21}$$

Here Z_A denotes the atomic number of nucleus A, m_e is the electron mass and M_A is the mass of nucleus A. As can be seen, the above Hamiltonian contains two kinetic energy terms, one for the nuclei and one for the electrons and the electron degrees of freedom are coupled to the degrees of freedom belonging to the nuclei.

At this point, it becomes convenient to introduce a new system of natural units, where the electron mass m_e, the electron charge e, the reduced Planck's constant $\hbar$ and the Coulomb constant $1/4\pi\epsilon_0$ are all chosen to be unity. Then the above Hamiltonian reduces to

$$\begin{aligned}\hat{H} &= -\sum_i \frac{1}{2}\nabla_i^2 - \sum_A \frac{1}{2M_A}\nabla_A^2 - \sum_{i,A} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &\quad + \sum_{A>B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}\end{aligned} \tag{2.22}$$

where M_A is now measured in units of m_e, energy is measured in units of $\alpha^2 m_e c^2$ called hartree units (where α is the fine structure constant) and length is measured in units of $a_0 = \hbar/(m_e c\alpha)$ (where a_0 is known as the Bohr radius).

2.7 The Born–Oppenheimer approximation

The way the nuclear and electronic degrees of freedom are coupled in the molecular Schrödinger equation makes it difficult to solve for the eigenfunctions of the time-independent Hamiltonian. A commonly made approximation, named after Born and Oppenheimer [14], is therefore used to decouple the motion of the nuclei from that of the electrons. The step is usually justified by considering that the forces acting on both electrons and nuclei are equal and opposite, meaning that their momenta must be similar. From this it follows that because the mass of the nucleus is much larger (of the order of 10^3 for Hydrogen atoms to 10^5 for heavy elements) than that of the electron, the velocity of the nuclei must be much slower than that of the electrons. This idea is used to justify the following Ansatz for the eigenstate of the full Hamiltonian [15]:

$$\Psi(\mathbf{r}, \mathbf{R}) = \chi(\mathbf{r}; \mathbf{R})\phi(\mathbf{R}). \tag{2.23}$$

Here, $\chi(\mathbf{r};\mathbf{R})$ is taken to depend only parametrically on the nucleic positions $\mathbf{R}$.

One now splits up the full Hamiltonian $\hat{H}$ into $\hat{H} = \hat{T}_n + \hat{H}_e(\mathbf{R})$, the nucleic kinetic energy term and the effective electron Hamiltonian where

$$\hat{H}_e(\mathbf{R}) = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_{i,A} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{A>B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \tag{2.24}$$

and requires $\chi(\mathbf{r};\mathbf{R})$ to be an eigenfunction of $\hat{H}_e(\mathbf{R})$ such that

$$\hat{H}_e(\mathbf{R})\chi(\mathbf{r};\mathbf{R}) = E_e(\mathbf{R})\chi(\mathbf{r};\mathbf{R}). \tag{2.25}$$

Note that the energy eigenvalues of the electronic part of the system depend on the nucleic positions $\mathbf{R}$. The quantity $E_e(\mathbf{R})$ is thus often referred to as the Born–Oppenheimer potential energy surface of the system.

Acting with the full Hamiltonian $\hat{H}$ on the Ansatz for the wavefunction of equation (2.23), one obtains

$$\begin{aligned} \hat{H}\Psi &= \chi(\mathbf{r};\mathbf{R}) \left[-\sum_A \frac{1}{2M_A}\nabla_A^2 + E_e(\mathbf{R}) \right] \phi(\mathbf{R}) \\ &\quad - \sum_A \frac{1}{2M_A} \left[2\nabla_A \chi(\mathbf{r};\mathbf{R}) \cdot \nabla_A \phi(\mathbf{R}) + \phi(\mathbf{R})\nabla_A^2 \chi(\mathbf{r};\mathbf{R}) \right]. \end{aligned} \tag{2.26}$$

Ignoring the second term of the above expression, ie. assuming that the terms involving the derivative of $\chi(\mathbf{r};\mathbf{R})$ with respect to the nucleic positions $\mathbf{R}$ are small, leads to the Born–Oppenheimer approximation, where the nucleic and electronic degrees of freedom are decoupled and $\phi(\mathbf{R})$ can be simply obtained from

$$\left[-\sum_A \frac{1}{2M_A}\nabla_A^2 + E_e(\mathbf{R}) \right] \phi(\mathbf{R}) = E\phi(\mathbf{R}) \tag{2.27}$$

The above equation is a Schrödinger-like equation for the ions moving in the effective potential $E_e(\mathbf{R})$. Thus the nuclei move on the potential energy surface defined by $\hat{H}_e$.

It can be shown [15] that the first term involving derivatives of $\chi(\mathbf{r};\mathbf{R})$ with respect to $\mathbf{R}$ in equation (2.26) strictly vanishes, while the second term is indeed of the order of $1/M_A$, thus justifying the approximation.

2.8 Perturbation theory

Given the complicated nature of many quantum mechanical systems, it is often beneficial to describe the complicated system via a simpler system that can be solved exactly. The

difference between the Hamiltonian of the complicated system and the Hamiltonian of the simple system is referred to as a perturbation. In case the perturbation is weak enough, the properties of the full system can be represented by an asymptotic series of corrections added to the solutions of the simple system.

Consider some system with a Hamiltonian $\hat{H}_0$ for which all (non-degenerate) eigenstates and energies are known, ie.

$$\hat{H}_0|\psi_n^{\{0\}}\rangle = E_n^{\{0\}}|\psi_n^{\{0\}}\rangle. \tag{2.28}$$

The full Hamiltonian $\hat{H}$ is now considered in terms of a perturbation to $\hat{H}_0$:

$$\hat{H} = \hat{H}_0 + \lambda V \tag{2.29}$$

where V is a weak perturbing potential and λ is a dimensionless continuous parameter that varies from 0 to 1, with 0 representing no perturbation and 1 representing a full perturbation. If the perturbation is weak, then the eigenstates and energies of the full system can be written as

$$E_n = \lambda^0 E_n^{\{0\}} + \lambda^1 E_n^{\{1\}} + \lambda^2 E_n^{\{2\}} + ... \tag{2.30}$$

$$|\psi_n\rangle = \lambda^0|\psi_n^{\{0\}}\rangle + \lambda^1|\psi_n^{\{1\}}\rangle + \lambda^2|\psi_n^{\{2\}}\rangle + ... \tag{2.31}$$

In the limit of weak perturbations, the above expansion should converge rapidly such that only the first couple of correction terms need to be computed in order to obtain a good approximation to the perturbed system.

The perturbation expansion of $|\psi_n\rangle$ can then be treated as an Ansatz acting on the perturbed Hamiltonian. To calculate the first order correction to the energy and wavefunction one can then ignore all terms containing higher orders of λ. To first order, one obtains

$$\lambda^0\hat{H}_0|\psi_n^{\{0\}}\rangle + \lambda^1\left(\hat{H}_0|\psi_n^{\{1\}}\rangle + V|\psi_n^{\{0\}}\rangle\right) = \lambda^0 E_n^{\{0\}}|\psi_n^{\{0\}}\rangle + \lambda^1\left(E_n^{\{0\}}|\psi_n^{\{1\}}\rangle + E_n^{\{1\}}|\psi_n^{\{0\}}\rangle\right) \tag{2.32}$$

The zeroth order of λ simply gives the unperturbed Schrödinger equation. The first order correction is then given by

$$\hat{H}_0|\psi_n^{\{1\}}\rangle + V|\psi_n^{\{0\}}\rangle = E_n^{\{0\}}|\psi_n^{\{1\}}\rangle + E_n^{\{1\}}|\psi_n^{\{0\}}\rangle. \tag{2.33}$$

Operating with $\langle\psi_n^{\{0\}}|$ on both sides one obtains

$$E_n^{\{0\}}\langle\psi_n^{\{0\}}|\psi_n^{\{1\}}\rangle + \langle\psi_n^{\{0\}}|V|\psi_n^{\{0\}}\rangle = E_n^{\{0\}}\langle\psi_n^{\{0\}}|\psi_n^{\{1\}}\rangle + E_n^{\{1\}}\langle\psi_n^{\{0\}}|\psi_n^{\{0\}}\rangle \quad (2.34)$$

$$\Rightarrow \langle\psi_n^{\{0\}}|V|\psi_n^{\{0\}}\rangle = E_n^{\{1\}}. \quad (2.35)$$

Here, the fact is used that $\{\psi_n^{\{0\}}\}$ are normalised. Thus, the first order correction to the energy eigenvalues is simply the expectation value of the perturbing potential with respect to the unperturbed eigenstates.

To compute correction to the unperturbed eigenstates, consider $V|\psi_n^{\{0\}}\rangle$:

$$\begin{aligned} V|\psi_n^{\{0\}}\rangle &= \hat{I}V|\psi_n^{\{0\}}\rangle = \sum_m |\psi_m^{\{0\}}\rangle\langle\psi_m^{\{0\}}|V|\psi_n^{\{0\}}\rangle \\ &= \left(\sum_{m\neq n} |\psi_m^{\{0\}}\rangle\langle\psi_m^{\{0\}}|\right) V|\psi_n^{\{0\}}\rangle + |\psi_n^{\{0\}}\rangle\langle\psi_n^{\{0\}}|V|\psi_n^{\{0\}}\rangle \\ &= \left(\sum_{m\neq n} |\psi_m^{\{0\}}\rangle\langle\psi_m^{\{0\}}|\right) V|\psi_n^{\{0\}}\rangle + E_n^{\{1\}}|\psi_n^{\{0\}}\rangle \end{aligned} \quad (2.36)$$

Here, fact that $\{\psi_n^{\{0\}}\}$ froms a complete set is used. Now substituting from equation (2.33) for $E_n^{\{1\}}|\psi_n^{\{0\}}\rangle$ and rearranging, one obtains

$$\left(E_n^{\{0\}} - \hat{H}_0\right)|\psi_n^{\{1\}}\rangle = \left(\sum_{m\neq n} |\psi_m^{\{0\}}\rangle\langle\psi_m^{\{0\}}|\right) V|\psi_n^{\{0\}}\rangle \quad (2.37)$$

and after operating with $\langle\psi_{m'}^{\{0\}}|$ on both sides and again using the orthogonality of the eigenstates, the first order correction to the wavefunctions is found to be

$$|\psi_n^{\{1\}}\rangle = \sum_{m\neq n} \frac{\langle\psi_m^{\{0\}}|V|\psi_n^{\{0\}}\rangle}{E_n^{\{0\}} - E_m^{\{0\}}}|\psi_m^{\{0\}}\rangle \quad (2.38)$$

where it is again assumed that the eigenstates of $\hat{H}_0$ are non-degenerate.

Furthermore, $|\psi_n\rangle$ is required to be normalised. Expanding the normalisation condition to the first order in λ yields

$$\langle\psi_n|\psi_n\rangle = \lambda^0\langle\psi_n^{\{0\}}|\psi_n^{\{0\}}\rangle + \lambda^1\left(\langle\psi_n^{\{1\}}|\psi_n^{\{0\}}\rangle + \langle\psi_n^{\{0\}}|\psi_n^{\{1\}}\rangle\right) = 1 \quad (2.39)$$

$$\Rightarrow \langle\psi_n^{\{1\}}|\psi_n^{\{0\}}\rangle = -\langle\psi_n^{\{0\}}|\psi_n^{\{1\}}\rangle = 0. \quad (2.40)$$

The power of the perturbative treatment lies in the fact that the first order changes to the eigenstates and eigenenergies only require knowledge of the zeroth order eigenstates

and eigenenergies. The same remains true for higher order perturbations, with the second order energy correction being equal to

$$E_n^{\{2\}} = \sum_{m \neq n} \frac{|\langle \psi_m^{\{0\}} | V | \psi_n^{\{0\}} \rangle|^2}{E_n^{\{0\}} - E_m^{\{0\}}}. \tag{2.41}$$

2.9 The Hartree–Fock method

While the Born–Oppenheimer approximation is a great simplification to the molecular Hamiltonian introduced in 2.6, it still requires the treatment of the electronic Hamiltonian $\hat{H}_e$ which cannot be solved analytically for its eigenstates and eigenvalues for anything but the most trivial systems. Indeed, even solving the time-independent Schrödinger equation for $\hat{H}_e$ numerically requires a computational effort that scales exponentially with the number of electrons. The reason for this lies in the Coulomb interaction between individual electrons. The interaction couples the degrees of freedom of all electrons and makes it impossible to separate them. In practice, approximate methods need to be used in order to solve for the eigenstates of $\hat{H}_e$. The simplest of these methods originates from restricting the state vector Ψ_e to a certain form.

Consider a system of N electrons where the task is to find an approximation to the many-electron wavefunction $\Psi_e(\mathbf{x}_1, \mathbf{x}_2 \cdots, \mathbf{x}_N)$. The easiest possible form that can be considered is

$$\Psi_e(\mathbf{x}_1, \cdots \mathbf{x}_N) = \psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2) \cdots \psi_N(\mathbf{x}_N) \tag{2.42}$$

where it is assumed that $\{\psi_i\}$ is an orthonormal set of single-electron eigenstates. However, this model form violates the anti-symmetry requirement of the wavefunction since

$$\begin{aligned} &\Psi_e(\cdots, \mathbf{x}_i, \cdots \mathbf{x}_j, \cdots) = -\Psi_e(\cdots, \mathbf{x}_j, \cdots \mathbf{x}_i, \cdots) \\ &\text{iff } \psi_i(\mathbf{x}_i)\psi_j(\mathbf{x}_j) = -\psi_i(\mathbf{x}_j)\psi_j(\mathbf{x}_i) \end{aligned} \tag{2.43}$$

which clearly is not generally the case. The simplest way of generating a guaranteed antisymmetric wavefunction from single-particle states ψ_i is in the form of a Slater-determinant [16]:

$$\Psi_e(\mathbf{x}_1, \cdots \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix} \tag{2.44}$$

Given the chosen Ansatz for the many-electron wavefunction in equation (2.44), one

has to work out the expectation value of the Hamiltonian which is straightforward due to the fact Ψ_e is made up from single electron state vectors:

$$\langle\Psi_e|\hat{H}_e|\Psi_e\rangle = -\sum_i^N \sum_\sigma \int \mathrm{d}^3r\, \frac{1}{2}\psi_i^*(\mathbf{x})\nabla^2\psi_i(\mathbf{x}) + \sum_{A>B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} + E_{\mathrm{eI}} + E_{\mathrm{ee}} \tag{2.45}$$

Here, E_{eI} is the energy due to electron-ion interaction and E_{ee} denotes electron-electron Coulomb interaction energy given by:

$$E_{\mathrm{ee}} = \frac{1}{2}\sum_i \sum_{j\neq i} \sum_{\sigma,\sigma'} \int \mathrm{d}^3r\, \mathrm{d}^3r' \left[\frac{|\psi_i(\mathbf{x})|^2|\psi_j(\mathbf{x}')|^2}{|\mathbf{r}-\mathbf{r}'|} - \frac{\psi_i^*(\mathbf{x})\psi_j(\mathbf{x}')\psi_i(\mathbf{x}')\psi_j^*(\mathbf{x})}{|\mathbf{r}-\mathbf{r}'|}\right]. \tag{2.46}$$

The first term of the above expression is known as the Hartree energy, which is the energy contribution due to the mean-field electron repulsion. The second term is the exchange energy and is purely a result of the anti-symmetry requirement of the all-electron wavefunction. Considering that for the term in the sum for which $i = j$, the Hartree contribution exactly cancels the exchange contribution, the $i \neq j$ condition in the double sum over states can be dropped.

Given the expectation value for the Hamiltonian, one can now make use the variational principle to find the lowest energy eigenstate of the Hamiltonian $\hat{H}_e$:

$$\begin{aligned} & \frac{\delta}{\delta\Psi_e}\frac{\langle\Psi_e|\hat{H}_e|\Psi_e\rangle}{\langle\Psi_e|\Psi_e\rangle} = 0 \\ \Rightarrow \quad & \frac{\delta}{\delta\Psi_e}\langle\Psi_e|\hat{H}_e|\Psi_e\rangle - E_0\frac{\delta}{\delta\Psi_e}\langle\Psi_e|\Psi_e\rangle = 0 \\ \Rightarrow \quad & \sum_i \frac{\delta}{\delta\psi_i}\langle\Psi_e|\hat{H}_e|\Psi_e\rangle = \sum_i \frac{\partial}{\partial\psi_i}\lambda_{ij}\left(\langle\psi_i|\psi_j\rangle - \delta_{ij}\right) \end{aligned} \tag{2.47}$$

where for the last line, the variations in the single-particle orbitals are considered and the Lagrange multiplier λ_{ij} is introduced to enforce the orthonormality between all single particle orbitals $\{\psi_i\}$ during the variation. From this expression it is possible to derive the N coupled Hartree-Fock equations for the single-particle orbitals:

$$\begin{aligned} & \left[-\frac{1}{2}\nabla^2 + \sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|}\right]\psi_i(\mathbf{x}) + \sum_{\sigma'}\left[\int \mathrm{d}^3r'\, \frac{\sum_j |\psi_j(\mathbf{x}')|^2}{|\mathbf{r}-\mathbf{r}'|}\right]\psi_i(\mathbf{x}) \\ & -\sum_j\sum_{\sigma'}\left[\int \mathrm{d}^3r'\, \frac{\psi_j^*(\mathbf{x}')\psi_i(\mathbf{x}')}{|\mathbf{r}-\mathbf{r}'|}\right]\psi_j(\mathbf{x}) = \epsilon_i\psi_i(\mathbf{x}) \end{aligned} \tag{2.48}$$

where ϵ_i is the effective single-particle energy associated with single-particle state ψ_i. Since the effective potential in this Schrödinger-like equation for each of the single particle states

ψ_i depends on the whole set $\{\psi_i\}$, the set of equations has to be solved self-consistently. To do so, one starts with a guess for the set of single-particle states, which is used to generate the effective potential for each of the single particle states. Solving equation (2.48) then leads to a new set $\{\psi_i\}$. The process can then be repeated until self-consistency is reached. The resulting Hartree–Fock energy is, by the variational principle, higher than the true ground-state energy of the system. The energy difference between the Hartree-Fock energy and the true energy is labelled "correlation energy" because it is due to correlation effects between the electrons that is ignored in the mean-field Hartree–Fock treatment of the electron-electron interaction[4].

The Hartree–Fock method is the first and oldest of the "practical methods" that are considered in this work and, by the variational principle, necessarily yields the best wavefunction that can be constructed from a single Slater determinant. Extensions to the Hartree-Fock method either introduce correlation into the theory via perturbation theory (Møller-Plesset perturbation theory) [17] or by extending the naive wavefunction Ansatz, either by using an exponential Ansatz of Slater determinants (Coupled Cluster) [18] or by considering multiple reference configurations (see [19] for an overview over so-called multireference approaches). In contrast to the methods that are discussed in the following chapters, all these methods are wavefunction methods, since they directly obtain an approximation for the ground state many-electron wavefunction of the quantum system.

[4]It should be noted here that the above description of "correlation energy" assumes the infinite basis set limit, where a complete basis set is used to represent the single-particle orbitals and to compute the Hartree–Fock energy. Furthermore, the term "true ground-state energy" does refer to the ground-state energy of the electron Hamiltonian $\hat{H}_e$ and thus all effects that are ignored in $\hat{H}_e$, like any relativistic effects and explicit spin dependence of the Hamiltonian do not contribute to the ground-state energy.

Chapter 3

Approximations to the ground state

In this chapter, approximations to calculating the ground state of a many-body system with electronic Hamiltonian $\hat{H}_e$ as discussed in the previous section are considered. Unlike the methods introduced briefly in the last chapter, the method discussed here avoids any explicit representation of the many-electron wavefunction, but will instead be making use of the electron density of the system defined via

$$\rho(\mathbf{r}) = \langle\Psi|\hat{\rho}(\mathbf{r})|\Psi\rangle = \int \prod_{i=2} \mathrm{d}^3 r_i \, |\Psi(\mathbf{r}, \mathbf{r}_2, \cdots \mathbf{r}_N)|^2. \tag{3.1}$$

One can readily see the advantage of such a treatment. While the many-body wavefunction is a $3N$-dimensional function, the electron density is a 3-dimensional function and thus potentially much easier to treat. For convenience, the spin dependence of the wavefunction is dropped in the entire chapter and only reintroduced where necessary.

3.1 Density-functional theory

In this section the main theorems that lead to the establishment of density-functional theory (DFT) are discussed. The electron Hamiltonian $\hat{H}_e$ of an interacting many-electron system is considered under the assumption that the ground state of $\hat{H}_e$ is non-degenerate. The Hamiltonian can be divided into three parts (ignoring the constant term due to ion-ion interactions that does not influence the solutions to the eigenstates):

$$\hat{H}_e = \hat{T} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{ext}} \tag{3.2}$$

where

$$\hat{V}_{\mathrm{ext}} = \sum_i V_{\mathrm{ext}}(\mathbf{r}_i) = -\sum_{i,A} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} \tag{3.3}$$

the static external potential due to the ions acting on the electrons and $\hat{T}$ and $\hat{V}_{\mathrm{ee}}$ are the kinetic energy term and the electron-electron interaction. Note that $\hat{T} + \hat{V}_{\mathrm{ee}}$ is the same

for any N-electron system. Therefore $\hat{H}_e$ and thus $|\Psi_0\rangle$ must be completely determined by N and $V_{\text{ext}}(\mathbf{r})$. Furthermore, because $|\Psi_0\rangle$ is a functional of N and $V_{\text{ext}}(\mathbf{r})$, the ground state density of the system, $\rho_0(\mathbf{r})$, is equally a functional of N and $V_{\text{ext}}(\mathbf{r})$.

3.1.1 Hohenberg–Kohn theorem

In 1964, Hohenberg and Kohn [7] proved two crucial theorems that form the starting point of density-functional theory:

- There exists a one-to-one mapping between the external potential $V_{\text{ext}}(\mathbf{r})$ of the interacting quantum system and the ground state density $\rho_0(\mathbf{r})$.
- For all v-representable[5] densities $\rho(\mathbf{r})$, there exists a unique functional $E[\rho(\mathbf{r})]$ such that the exact ground state energy is the global minimum of $E[\rho(\mathbf{r})]$.

The first Hohenberg–Kohn theorem can be proven by assuming that there exists a second external potential $V'_{\text{ext}}(\mathbf{r})$ with ground state wavefunction $|\Psi'_0\rangle$ that produces the same ground state density $\rho_0(\mathbf{r})$. One therefore considers two different Hamiltonians $\hat{H}_e = \hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\text{ext}}$ and $\hat{H}'_e = \hat{T} + \hat{V}_{\text{ee}} + \hat{V}'_{\text{ext}}$ with ground state energies $E_0 = \langle\Psi_0|\hat{H}_e|\Psi_0\rangle$ and $E'_0 = \langle\Psi'_0|\hat{H}'_e|\Psi'_0\rangle$. Using the variational principle it is clear that

$$E_0 < \langle\Psi'_0|\hat{H}_e|\Psi'_0\rangle = \langle\Psi'_0|\hat{H}'_e|\Psi'_0\rangle + \langle\Psi'_0|(\hat{H}_e - \hat{H}'_e)|\Psi'_0\rangle \tag{3.4}$$

$$E'_0 < \langle\Psi_0|\hat{H}'_e|\Psi_0\rangle = \langle\Psi_0|\hat{H}_e|\Psi_0\rangle + \langle\Psi_0|(\hat{H}'_e - \hat{H}_e)|\Psi_0\rangle \tag{3.5}$$

Given that both systems have the same ground state density, the terms $\langle\Psi'_0|(\hat{H}_e - \hat{H}'_e)|\Psi'_0\rangle$ and $\langle\Psi_0|(\hat{H}'_e - \hat{H}_e)|\Psi_0\rangle$ can both be written in terms of $\rho_0(\mathbf{r})$ to give

$$E_0 < E'_0 + \int \rho_0(\mathbf{r}) \left[V_{\text{ext}}(\mathbf{r}) - V'_{\text{ext}}(\mathbf{r})\right] \mathrm{d}^3 r \tag{3.6}$$

$$E'_0 < E_0 - \int \rho_0(\mathbf{r}) \left[V_{\text{ext}}(\mathbf{r}) - V'_{\text{ext}}(\mathbf{r})\right] \mathrm{d}^3 r \tag{3.7}$$

from which immediately follows that

$$E_0 + E'_0 < E'_0 + E_0 \tag{3.8}$$

which is a contradiction. Thus if two systems have the same ground state density, their external potentials must be identical (to within a constant) and vice versa. Therefore, there exists a one-to-one mapping between the external potential and the ground state density of the system.

[5] A density $\rho_\alpha(\mathbf{r})$ is v-representable, iff it is the ground state density of a system with some external potential $V^\alpha_{\text{ext}}(\mathbf{r})$. See later in this section for more details on the question of v-representability.

To consider the second theorem, it is first necessary to address some of the implications of the first theorem. Since the ground state density uniquely determines the external potential and the external potential uniquely defines the electronic Hamiltonian of the system, it must be possible to express any ground state property of the many-body system as a unique functional of the ground state density. Now consider a set of all densities $\{\rho_v\}$, each of which is the ground state density to some external potential $V_v(\mathbf{r})$ and thus $\{\rho_v\}$ uniquely defines the set of many-body wavefunctions $\{|\Psi_v\rangle\}$. These densities are referred to as v-representable since for them the one-to-one mapping to the potential exists. Since the wavefunction is uniquely defined for the v-representable density, the functional $F[\rho_v] = \langle\Psi_v|\hat{T} + \hat{V}_{\text{ee}}|\Psi_v\rangle$ must be uniquely defined as well. Thus for the v-representable densities there exists a unique energy functional $E[\rho_v]$ such that

$$\begin{aligned} E[\rho_v] &= \langle\Psi_v|\hat{H}_e|\Psi_v\rangle = \langle\Psi_v|\hat{T} + \hat{V}_{\text{ee}}|\Psi_v\rangle + \langle\Psi_v|\hat{V}_{\text{ext}}|\Psi_v\rangle \\ &= F[\rho_v] + \int \mathrm{d}^3r\, V_{\text{ext}}(\mathbf{r})\rho_v(\mathbf{r}) \end{aligned} \tag{3.9}$$

for some electronic Hamiltonian $\hat{H}_e$ with external potential $\hat{V}_{\text{ext}}$.

Now consider some v-representable density ρ_v with associated potential $V_v(\mathbf{r})$ and wavefunction $|\Psi_v\rangle$ acting on the Hamiltonian $\hat{H}_e$ with arbitrary external potential V_{ext}. Using both the variational principle and the considerations above one finds that

$$\langle\Psi_v|\hat{H}_e|\Psi_v\rangle = F[\rho_v] + \int \mathrm{d}^3r\, V_{\text{ext}}(\mathbf{r})\rho_v(\mathbf{r}) = E[\rho_v] \geq E_0 \tag{3.10}$$

where equality is only reached if $\rho_v = \rho_0$. Thus, in principle it is possible to find the ground state energy of any system defined by the electronic Hamiltonian $\hat{H}_e$ by minimising the functional $E[\rho_v]$ over the set of all v-representable densities $\{\rho_v\}$. However the Hohenberg–Kohn theorems do not specify how the minimisation should be carried out in practice given that, for an arbitrary density, it is generally not known if it is v-representable.

3.1.2 Levy constrained search method

The problem of carrying out the minimisation of the energy functional suggested by Hohenberg and Kohn without relying on v-representability has been addressed by Levy [20]. It turns out that to prove the second statement of the Hohenberg–Kohn theorem, it is not necessary to make a reference to v-representability at all. Following Levy it is possible to define the functional $F[\rho]$ as

$$F[\rho] = \min_{|\Psi\rangle\to\rho} \langle\Psi|\hat{T} + \hat{V}_{\text{ee}}|\Psi\rangle. \tag{3.11}$$

Thus $F[\rho]$ is defined via the minimum of the expectation value of $\left(\hat{T}+\hat{V}_{\text{ee}}\right)$, where the minimisation is carried out with respect to all many-electron wavefunctions $|\Psi\rangle$ that yield the electron density ρ.

Now consider the many-electron wavefunction $|\Psi_{[\rho]}\rangle$ that produces the density ρ and minimizes $F[\rho]$. The expectation value of the Hamiltonian $\hat{H}_e$ with external potential V_{ext} and ground state energy E_0, under the redefinition of $E[\rho]$ via the new definition of $F[\rho]$ can then be written as

$$\begin{aligned}\langle\Psi_{[\rho]}|\hat{H}_e|\Psi_{[\rho]}\rangle &= \langle\Psi_{[\rho]}|\hat{T}+\hat{V}_{\text{ee}}|\Psi_{[\rho]}\rangle + \langle\Psi_{[\rho]}|\hat{V}_{\text{ext}}|\Psi_{[\rho]}\rangle \\ &= F[\rho] + \int \mathrm{d}^3r\, V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) = E[\rho] \geq E_0 \end{aligned} \tag{3.12}$$

where the last statement follows necessarily from the variational principle and equality only holds if $|\Psi_{[\rho]}\rangle = |\Psi_0\rangle$. The statement holds for all densities that can be generated by an N-particle wavefunction, a condition that is referred to as N-representability and that is much less restrictive than v-representability. Now consider $F[\rho_0]$ and the ground-state wavefunction $|\Psi_0\rangle$. Clearly $|\Psi_0\rangle$ is one of the many-electron wavefunctions yielding ρ_0 over which $F[\rho]$ is minimised. From equation (3.11) it therefore follows that

$$\begin{aligned} & F[\rho_0] \leq \langle\Psi_0|\hat{T}+\hat{V}_{\text{ee}}|\Psi_0\rangle \\ \Rightarrow \quad & F[\rho_0] + \int \mathrm{d}^3r\, V_{\text{ext}}(\mathbf{r})\rho_0(\mathbf{r}) \leq \langle\Psi_0|\hat{H}_e|\Psi_0\rangle \\ \Rightarrow \quad & E[\rho_0] \leq E_0 \end{aligned} \tag{3.13}$$

From the two inequalities, it directly follows that $E[\rho] \geq E[\rho_0] = E_0$ such that minimising $E[\rho]$ with respect to all ρ that are N-representable yields the ground state energy of the system. It has been shown [21, 22] that the condition of N-representability is easily met for arbitrary, normalisable, differentiable densities and that it poses no real restriction on the validity of the central Hohenberg–Kohn theorem as the problem of v-representability does.

3.1.3 Kohn–Sham theory

While the Hohenberg–Kohn theorems prove the existence of a functional $E[\rho]$ that, if minimised, yields the ground state density and energy of the many-electron system, the exact form of $F[\rho]$ is unknown and has to be approximated.

The main strategy to achieve approximations is to split $F[\rho]$ into independent particle and mean field contributions and unknown many-body contributions. One therefore writes

$$F[\rho] = T_s[\rho] + E_H[\rho] + E_{\text{xc}}[\rho] \tag{3.14}$$

where $T_s[\rho]$ is the kinetic energy of a system of N non-interacting particles of density ρ and $E_H[\rho]$ is the Hartree energy (see section 2.9), the mean-field electron interaction energy due to a system of N particles with density ρ. E_{xc} is the exchange-correlation energy and contains all many-body effects not included in $T_s[\rho]$ and $E_H[\rho]$, such that

$$E_{\mathrm{xc}}[\rho] = T[\rho] - T_s[\rho] + E_{\mathrm{ee}}[\rho] - E_H[\rho] \tag{3.15}$$

Since this is just a simple rewriting of the original statement, both $E_{\mathrm{xc}}[\rho]$ and $T_s[\rho]$ are generally not known and suitable approximations have to be found. It is to be hoped that the many-body corrections introduced through E_{xc} are small. However, the kinetic energy contributions through $T_s[\rho]$ are generally not small and good approximate treatments become crucial. Several approximations have been known long before the development of density-functional theory, most notably the Thomas–Fermi kinetic energy functional [23, 24]

$$T_{TF}[\rho] = \frac{3}{10}\left(3\pi^2\right)^{\frac{2}{3}} \int \mathrm{d}^3r \; [\rho(\mathbf{r})]^{\frac{5}{3}} \tag{3.16}$$

and the von Weizsäcker kinetic energy functional [25]

$$T_W[\rho] = \frac{3}{10}\left(3\pi^2\right)^{\frac{2}{3}} \int \mathrm{d}^3r \; [\rho(\mathbf{r})]^{\frac{5}{3}} + \frac{1}{8}\int \mathrm{d}^3r \, \frac{\nabla\rho(\mathbf{r})\cdot\nabla\rho(\mathbf{r})}{\rho(\mathbf{r})}. \tag{3.17}$$

While in the Thomas–Fermi model, the kinetic energy per infinitesimal unit volume is constant and only dependent on the density at that point, the von Weizsäcker functional adds a correction term depending on the gradient of $\rho(\mathbf{r})$. It can be shown that the von Weizsäcker functional is exact for 1- or 2-electron systems [26] and constitutes a lower bound to $T_s[\rho]$ for an N-electron system. Other, higher order corrections to $T_W[\rho]$ can be introduced but in general the deviation of these approximate kinetic energy functionals from the true, independent particle kinetic energy functional $T_s[\rho]$ is significant [26].

A way to avoid any approximations to $T_s[\rho]$ was introduced by Kohn and Sham [27]. It consists of a remapping of the problem of N interacting electrons to a fictitious system of non-interacting electrons in an effective potential producing the same ground-state density. It exploits the fact that for a set of N non-interacting electrons with single-particle wavefunctions $\{\psi_i(\mathbf{r})\}$ producing the density ρ, $T_s[\rho]$ is simply given by

$$T_s[\rho] = T_s[\{\psi_i[\rho]\}] = -\sum_i^{N/2} \langle\psi_i|\nabla^2|\psi_i\rangle \tag{3.18}$$

and is thus exactly known (here it is assumed that the system in question is spin-degenerate and thus every single-particle state ψ_i is doubly occupied). In order to arrive

at the Kohn–Sham equations, it is possible to minimize the energy functional of a system in external potential $V_{\text{ext}}(\mathbf{r})$. The stationarity condition at the minimum can be written as:

$$\frac{\delta}{\delta\rho(\mathbf{r})}\left[T_s[\rho] + E_H[\rho] + E_{\text{xc}}[\rho] + \int \mathrm{d}^3r\, V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) - \mu\left(\int \mathrm{d}^3r\, \rho(\mathbf{r}) - N\right)\right] = 0 \quad (3.19)$$

where the Lagrange multiplier μ is introduced to constrain the density to integrate to N. It follows that

$$\frac{\delta T_s[\rho]}{\delta\rho(\mathbf{r})} + V_{\text{KS}}(\mathbf{r}) = \mu \quad (3.20)$$

with

$$V_{\text{KS}}(\mathbf{r}) = \int \mathrm{d}^3r'\, \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E_{\text{xc}}[\rho]}{\delta\rho(\mathbf{r})} + V_{\text{ext}}(\mathbf{r}). \quad (3.21)$$

However, Eqn. (3.20) has the form of a noninteracting system of particles moving in an effective potential $V_{\text{KS}}(\mathbf{r})$. Thus the problem of finding the ground state density of the system is mapped to solving the single particle Schrödinger equation for a set of $N/2$ single-particle states $\psi_i(\mathbf{r})$ with

$$\left[-\frac{1}{2}\nabla^2 + \hat{V}_{\text{KS}}\right]|\psi_i\rangle = \epsilon_i|\psi_i\rangle. \quad (3.22)$$

The electron density of the system is then defined via the sum of the single particle densities

$$\rho(\mathbf{r}) = 2\sum_i^{N/2} |\psi_i(\mathbf{r})|^2 \quad (3.23)$$

and the total energy of the system can be written as

$$\begin{aligned} E[\rho] &= T_s[\{\psi_i[\rho]\}] + E_H[\rho] + E_{\text{xc}}[\rho] + \int \mathrm{d}^3r\, V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) \\ &= 2\sum_i^{N/2}\langle\psi_i|\hat{H}_{\text{KS}}|\psi_i\rangle - \frac{1}{2}E_H[\rho] - \int \mathrm{d}^3r\, \rho(\mathbf{r})V_{\text{xc}}(\mathbf{r}) + E_{\text{xc}}[\rho] \\ &= 2\sum_i^{N/2}\epsilon_i - \frac{1}{2}E_H[\rho] - \int \mathrm{d}^3r\, \rho(\mathbf{r})V_{\text{xc}}(\mathbf{r}) + E_{\text{xc}}[\rho] \quad (3.24) \\ &= E_{\text{BS}}[\rho] - E_{\text{dc}}[\rho]. \quad (3.25) \end{aligned}$$

Here, the effective Kohn–Sham Hamiltonian $\hat{H}_{\text{KS}}$ has been introduced and the ground state energy is written in terms of the bandstructure energy E_{BS} of single particle energy eigenvalues ϵ_i and some double counting terms that are obtained from comparing to the full energy functional of the density.

Since the Kohn–Sham effective potential in (3.22) is dependent on the density, which is in turn defined via the Kohn–Sham single particle orbitals $\{|\psi_i\rangle\}$, the problem has to be solved self-consistently. This is done by starting with a guess for the ground-state density of the system which defines the Kohn–Sham potential and a set $\{|\psi_i\rangle\}$. These orbitals define a new density and the process can be iterated until the change between incoming and outgoing density vanishes[6].

Thus in order to gain a good representation of the kinetic energy functional, the problem of minimising $E[\rho]$ with respect to ρ is recast into iteratively solving for $N/2$ orthogonal single particle states $|\psi_i\rangle$. This reformulation comes at a computational cost. While minimising $E[\rho]$ directly does scale linearly with size of the system, finding the solutions to (3.22) generally scales as $\mathcal{O}(N^3)$ with system size, due to the orthogonality constraints on the $N/2$ Kohn–Sham single-particle orbitals. However, while there has been an intense research effort over the last decades in trying to find better purely density-based approximations to $T_s[\rho]$ for orbital-free DFT approaches (OF-DFT) [26, 29], to this point none of them can achieve accuracies comparable to Kohn–Sham DFT over a wide range of different systems. For this reason Kohn–Sham DFT has been the more widely used flavour of DFT over the last decades.

While the Kohn–Sham eigenstates $|\psi_i\rangle$ are those of a fictitious noninteracting system in an effective potential that are introduced purely in order to gain a representation of the kinetic energy functional, they are often being taken to have some physical meaning attached to them. In particular, it is often found that Kohn–Sham eigenvalues are in reasonably good agreement with experimentally obtained bandstructures of the real system. Strictly speaking however, only the Kohn–Sham energy of the highest occupied orbital has a physical meaning attached to it, as for the exact DFT functional it corresponds to the ionisation energy of the system following Koopmans' theorem [30].

3.1.4 Exchange-correlation functionals

So far, the form of the functional $E_{\text{xc}}[\rho]$, the term in which all the many-body effects of exchange and correlation are collected that are ignored in $T_s[\{\psi_i[\rho]\}]$ and $E_H[\rho]$ has not been addressed. The exact functional form of $E_{\text{xc}}[\rho]$ is unknown and a number of properties that $E_{\text{xc}}[\rho]$ must obey suggest that it cannot be easily stated as a simple functional of the density [31]. Therefore in practice, DFT cannot provide an exact solution to the many-body problem and the results obtained are dependent on the way $E_{\text{xc}}[\rho]$ is approximated.

[6] In practice, a naive approach of using the outgoing density to define the Kohn–Sham potential is unlikely to yield convergence in real systems. For calculations in real systems, sophisticated density mixing schemes are used that take into account the densities of previous iterations to define a new one (see for example [28] for details).

The simplest possible approximation of $E_{\mathrm{xc}}[\rho]$ is the local density approximation (LDA), where it is assumed that the exchange-correlation energy at an infinitesimal volume element d^3r is the same as that of a homogeneous electron gas with uniform density $\rho = \rho(\mathbf{r})$. Thus

$$E_{\mathrm{xc}}[\rho] = \int \mathrm{d}^3r\, \epsilon_{\mathrm{xc}}\left[\rho(\mathbf{r})\right] \rho(\mathbf{r}) \tag{3.26}$$

$$V_{\mathrm{xc}}(\mathbf{r}) = \epsilon_{\mathrm{xc}}\left[\rho(\mathbf{r})\right] + \rho(\mathbf{r}) \left.\frac{\mathrm{d}\epsilon\left[\rho\right]}{\mathrm{d}\rho}\right|_{\rho=\rho(\mathbf{r})} \tag{3.27}$$

The exchange energy of the homogeneous electron gas is fully known, while for the correlation energy, analytic forms for the high and low density limit are known [32]. These limits are extrapolated over the whole density range by fitting to data obtained by very accurate quantum Monte-Carlo calculations [35]. Different fittings have yielded a number of different LDA functionals in recent years [32–34]. LDA is very appealing due to its simplicity and generally yields good results for a wide range of different materials and molecules.

While the LDA functionals have turned out to yield very good results in a variety of systems and have been used extensively by physicists, total energies calculated for molecular systems are often not good enough for the purposes of quantum chemists. As a result, DFT was not a widely used method in quantum chemistry until the next class of functionals, the gradient-corrected functionals (GGA functionals) were introduced [36]. In order to deal with rapidly changing electron densities such as encountered close to the nucleus[7], GGA functionals treat the exchange-correlation energy as

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = \int \mathrm{d}^3r\, \epsilon_{\mathrm{xc}}^{\mathrm{GGA}}\left[\rho(\mathbf{r}), |\nabla\rho(\mathbf{r})|\right] \rho(\mathbf{r}) \tag{3.28}$$

and different analytic forms and fittings to the homogeneous electron gas have again led to a vast variety of different functionals of this type [37–39]. GGA functionals do improve significantly on total energies and some structural properties like bond lengths in some covalently bonded systems, but in many regards have the same failures as LDA functionals and do not deliver a universal improvement over a wide range of sytems.

The main failure of many simple density-based exchange correlation functionals is the so-called self-interaction error. This error can be quantified by considering a system with a single electron. Since there is only one electron, the electron-electron interaction is

[7] This is mainly the case for quantum chemistry all-electron DFT methods, since they need to explicitly treat core electrons that are confined to the region close to the nucleus. These core electrons are ignored in pseudopotential approaches (see section 3.2.6) often used in solid-state physics community, making the electron density smoother.

required to strictly vanish. In the DFT framework, this means that the Hartree potential of a single electron is required to exactly cancel the exchange-correlation functional. In essence, the requirement states that the asymptotic $1/r$ dependence of the Hartree potential must be cancelled by a $1/r$ dependence of the exchange-correlation potential. For most (semi)-local functionals depending only locally on ρ (or $|\nabla\rho|$), this simple property is violated. Note that the Hartree–Fock method discussed in section 2.9 does not suffer from this type of self interaction error as the Hartree potential exactly cancels the exchange potential for a single electron.

The self-interaction error in most commonly used DFT functionals causes ionisation energies to be severely underestimated. This can be understood by considering that in the limit that an electron is moved far away from an arbitrary nucleus, it should be feeling a simple $1/r$ potential of the nucleus screened by all other valence electrons. However, in DFT treatments the self-interaction error screens the potential, causing it to decay faster, which leads to underestimated ionisation energies.

To address some of the shortcomings of (semi-)local functionals, a third class of functionals named hybrid functionals [40, 41] has been introduced. In hybrid functionals, a certain fraction of non-local Hartree–Fock exchange is added to a GGA-type exchange-correlation functional to yield better results in systems where the (semi-)local exchange correlation functionals perform poorly. However, while hybrid functionals often perform better in systems where non-local effects are important, they still do not produce the correct long-range behaviour for V_{xc} and are, unlike the pure Hartree–Fock method, not self-interaction free. They also come with significant increase in computational cost as V_{xc} is no longer a simple functional of the density but has a dependence on individual Kohn–Sham orbitals. Furthermore, just as (semi)-local functionals, hybrid functionals generally do not reproduce correct binding curves for systems that are bound by van der Waals type interactions which arise from non-local correlation effects[8].

3.2 From theory to practical method

In section 3.1 the theoretical foundation of density-functional theory was outlined. Over the last thirty years, intense research has been invested into transforming the theory into a practical method that can be used routinely for calculating energies and structural properties of any system. In this section, some of the main techniques used by most modern DFT-codes in order to facilitate routine calculations on a variety of systems are discussed.

[8]These nonlocal correlation effects can be introduced into the DFT framework by making use of van der Waals density functionals, which have been shown to give realistic binding curves in many systems where van der Waals effects are important (see [42])

3.2.1 Bloch's theorem, Brillouin zones and periodic boundary conditions

Consider a crystalline solid with a translational symmetry defined by the primitive lattice vectors $\mathbf{a}_i$. From this it follows that the external potential of the ions V_{ext} follows the same periodicity, ie.

$$V_{\text{ext}}(\mathbf{r} + n_i\mathbf{a}_i) = V_{\text{ext}}(\mathbf{r}) \text{ for } n_i \in \mathbb{Z},\, i \in \{1, 2, 3\} \tag{3.29}$$

It follows that

$$V_{\text{KS}}(\mathbf{r} + \mathbf{R}) = V_{\text{KS}}(\mathbf{r}) \tag{3.30}$$

where

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \text{ for } n_1, n_2, n_3 \in \mathbb{Z} \tag{3.31}$$

and thus that the effective single particle Hamiltonian $\hat{H}_{\text{KS}}$ possesses the same translational symmetry as the lattice since the kinetic energy is translation invariant. Consider now the translation operator $\hat{T}_{\mathbf{R}}$ that translates the system by $\mathbf{R}$. Since the Hamiltonian is invariant under translations by $\mathbf{R}$, $\hat{H}_{\text{KS}}$ and $\hat{T}_{\mathbf{R}}$ commute and therefore have a common set of eigenfunctions. From this consideration it immediately follows that $\psi_i^{\text{KS}}(\mathbf{r}+\mathbf{R})$ can be written as

$$\psi_i^{\text{KS}}(\mathbf{r} + \mathbf{R}) = C(\mathbf{R})\psi_i^{\text{KS}}(\mathbf{r}) \tag{3.32}$$

for some complex function $C(\mathbf{R})$. Given that translating the Kohn-Sham eigenfunction by $\mathbf{R}$ cannot change the normalisation of that wavefunction, it is found that $|C(\mathbf{R})|^2 = 1$ and that therefore

$$C(\mathbf{R}) = e^{\mathrm{i}\theta(\mathbf{R})}. \tag{3.33}$$

Considering that

$$\hat{T}_{\mathbf{R}}\hat{T}_{\mathbf{R}'}\psi_j^{\text{KS}}(\mathbf{r}) = \hat{T}_{\mathbf{R}'}\hat{T}_{\mathbf{R}}\psi_k^{\text{KS}}(\mathbf{r}) = \hat{T}_{\mathbf{R}+\mathbf{R}'}\psi_j^{\text{KS}}(\mathbf{r}) = \psi_j^{\text{KS}}(\mathbf{r} + \mathbf{R} + \mathbf{R}') \tag{3.34}$$

it follows that

$$\hat{T}_{\mathbf{R}}\psi_j^{\text{KS}}(\mathbf{r}) = \psi_j^{\text{KS}}(\mathbf{r} + \mathbf{R}) = e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\psi_j^{\text{KS}}(\mathbf{r}) \tag{3.35}$$

for some wavevector $\mathbf{k}$.

Equivalently one can state that the eigenfunctions of the Hamiltonian can be written as

$$\psi_{j\mathbf{k}}^{\text{KS}}(\mathbf{r}) = e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}u_j(\mathbf{r}) \tag{3.36}$$

where the function $u_i(\mathbf{r})$ has the same periodicity as the external potential. This can be

seen by considering $\psi^{\mathrm{KS}}_{j\mathbf{k}}(\mathbf{r}+\mathbf{R})$ and finding that

$$\psi^{\mathrm{KS}}_{j\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})}u_j(\mathbf{r}+\mathbf{R}) = e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}u_j(\mathbf{r}) = e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\psi^{\mathrm{KS}}_{j\mathbf{k}}(\mathbf{r}). \tag{3.37}$$

The two statements in (3.35) and (3.37) are known as Bloch's theorem [43].

Before continuing, one has to consider the appropriate boundary conditions for the problem at hand. It is assumed that the crystal considered here is made up of a number of N_j primitive unit cells in each of the three dimensions. An appropriate set of boundary conditions comes in the form of Born–von Karman boundary conditions that are imposed on the crystal supercell made up of N_j copies of the unit cell in all j direction. Let $\mathbf{o}$ denote the point of origin of the crystal, from which a number of N_j primitive unit cells get repeated in each of the three directions. One then requires that

$$\psi^{\mathrm{KS}}_i(\mathbf{o}+\sum_j^3 \mathbf{a}_j N_j) = \psi^{\mathrm{KS}}_i(\mathbf{o}) \tag{3.38}$$

To obey the boundary condition, it is required that $\mathbf{k}\cdot(\sum_j^3 \mathbf{a}_j N_j) = 2\pi l$ for some integer l. From this it simply follows that

$$\mathbf{k} = \sum_{j=1}^{3} \frac{m_j}{N_j}\mathbf{b}_j \quad \text{for} \quad N_j,\, m_j \in \mathbb{Z} \tag{3.39}$$

with the primitive reciprocal lattice vectors defined as

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}. \tag{3.40}$$

One can now use the above definitions to define an arbitrary reciprocal lattice vector $\mathbf{G}$ as

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 \quad \text{for} \quad m_1, m_2, m_3 \in \mathbb{Z}. \tag{3.41}$$

Note that from Bloch's theorem it follows than $\psi_{i\mathbf{k}}(\mathbf{r}) = \psi_{i[\mathbf{k}+\mathbf{G}]}(\mathbf{r})$ and thus it is sufficient to limit the range of allowed $\mathbf{k}$-vectors to $-\frac{1}{2} \leq \frac{m_j}{N_j} \leq \frac{1}{2}$ in order to obtain the full solution of the system. The reciprocal space to which $\mathbf{k}$ is limited by introducing this range to the allowed values of m_j is known as the first Brillouin zone and it contains all information of the entire system. From the above considerations it can be seen that the first Brillouin zone contains exactly $N_1 \times N_2 \times N_3$ discrete allowed values for $\mathbf{k}$, and thus there are as many $\mathbf{k}$-points to be considered as there are primitive unit cells in the crystal supercell.

To solve for $|\psi_{i\mathbf{k}}\rangle$ one can then use the second formulation of Bloch's theorem (3.37)

to act onto the Hamiltonian, which yields

$$\hat{H}^{\mathrm{KS}}_{\mathbf{k}}|u_{i\mathbf{k}}\rangle = \epsilon_i|u_{i\mathbf{k}}\rangle \tag{3.42}$$

under the boundary condition $u_{i\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{i\mathbf{k}}(\mathbf{r})$, where the effective Kohn–Sham Hamiltonian now has a $\mathbf{k}$-dependence that can be expressed as[9]

$$\hat{H}_{\mathbf{k}} = -\frac{1}{2}\left(\mathrm{i}\mathbf{k} + \nabla\right)^2 + V_{\mathrm{KS}}(\mathbf{r}). \tag{3.43}$$

An important special point in the Brillouin zone is the Γ-point corresponding to $\mathbf{k} = 0$. From the above equation (3.43) it can be seen that at this point, the eigenfunctions of the Hamiltonian can be chosen to be real.

Therefore, in order to calculate the properties of a crystal consisting of a number of unit cells, it is sufficient to solve (3.42) in a single unit cell for all allowed $\mathbf{k}$ in the first Brillouin zone. The total energy of the system is then defined via a discrete sum over eigenenergies in $\mathbf{k}$-space. In case of an infinite crystal, $N_i \to \infty$, $\mathbf{k}$ becomes continuous in the Brillouin zone and the Kohn–Sham eigenenergies form continuous bands in $\mathbf{k}$-space. For practical calculations however, one still only considers a discrete number of $\mathbf{k}$-states and the integral over the Brillouin zone is approximated as a discrete sum. If enough discrete points are used, the integral is well converged and the result obtained is equivalent to that of the infinite system.

From the above considerations it also follows than one does not have to choose the primitive unit cell in real space to solve for the Bloch functions $|u_{i\mathbf{k}}\rangle$. It is equally valid to consider larger real space cells made up of a number of primitive unit cells. While this increases the computational cost of solving for $\{|u_{i\mathbf{k}}\rangle\}$ it means that fewer discrete points $\mathbf{k}$ are required to achieve converged energies for the infinite crystal.

3.2.2 Wannier functions

Consider an infinite crystal with Bloch functions $|\psi_{i\mathbf{k}}\rangle$ as defined in (3.37). Let index i label an *isolated* Bloch function associated with an energy band $\epsilon_i(\mathbf{k})$ that does not possess any degeneracies with any other Bloch band for any value of $\mathbf{k}$. It is then possible to define a localised, orthonormal real space function $w_n(\mathbf{r}-\mathbf{R})$ such that [44]

$$|w_{n\mathbf{R}}\rangle = \frac{V}{(2\pi)^3}\int \mathrm{d}^3k\, e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}}|\psi_{i\mathbf{k}}\rangle \tag{3.44}$$

[9]Here, it is assumed that V_{KS} is fully local, which is not the case depending on whether non-local pseudopotentials are used in order to remove the core electrons from the effective Hamiltonian.

and, conversely

$$|\psi_{i\mathbf{k}}\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} |w_{n\mathbf{R}}\rangle \tag{3.45}$$

where V is the volume of the real space primitive unit cell. This function is referred to as a Wannier function. In a normal infinite crystal, there is generally some band crossing and thus degeneracies in $\mathbf{k}$-space. In this scenario one defines a set of J generalised Wannier functions for a set of J *composite* bands, that have some degeneracies amongst each other but not with any other band, via equation (3.44). This establishes a connection between generalised Wannier functions and generalised Bloch states via equation (3.45) such that the generalised Bloch states are no longer eigenfunctions of the Hamiltonian at $\mathbf{k}$ but are transformed to the eigenstates by a single unitary transformation defined by a matrix of size $J \times J$ [44]. Consider an insulator with some number of occupied bands N_{occ}. By definition, the set of N_{occ} bands forms a composite set of bands, not sharing any degeneracies in $\mathbf{k}$-space with unoccupied bands. It is therefore possible to define a set of N_{occ} generalised localised Wannier functions spanning the same space as the N_{occ} eigenstates $|\psi_{i\mathbf{k}}\rangle$.

There is some arbitrariness in defining Wannier functions which follows from the fact that there is a freedom in the choice of phase of Bloch functions $|u_{i\mathbf{k}}\rangle$. In practice this choice of freedom can be expressed as a unitary matrix of dimensions $N_{\text{occ}} \times N_{\text{occ}}$ for a composite set of N_{occ} bands of such that

$$|u_{i\mathbf{k}}\rangle \rightarrow \sum_{j=1}^{N_{\text{occ}}} U_{ji}^{\{\mathbf{k}\}} |u_{j\mathbf{k}}\rangle. \tag{3.46}$$

Given the arbitrariness of the phase of $|u_{i\mathbf{k}}\rangle$ it can be chosen such that the spread of the resulting Wannier functions is minimised, leading to maximally localised generalised Wannier functions. It is this feature of maximal localisation that makes Wannier functions interesting for a number of applications, including linear-scaling approaches as discussed in section 3.3.

The asymptotic localisation of Wannier functions has been discussed for a number of years [44]. While it had long been proven that Wannier functions and generalised Wannier functions are exponentially localised in 1D [45,46], in 3D the proof only existed for isolated bands [46,47]. Recently, it has been shown that generalised Wannier functions spanning the composite bands corresponding to the occupied states in a 3D insulator with time-reversal symmetry are exponentially localised [48].

3.2.3 Basis functions

The most effective way of solving the Kohn–Sham Hamiltonian for its eigenstates depends on the type of system of interest. In most cases, it is beneficial to represent $|\psi_i^{\mathrm{KS}}\rangle$ in some underlying set of basis functions $\{\phi_\alpha\}$ such that

$$|\psi_i^{\mathrm{KS}}\rangle = \sum_\alpha a_i^\alpha |\phi_\alpha\rangle \tag{3.47}$$

for some complex coefficients a_i^α. For the case of a periodic crystal discussed above, a natural choice are plane waves of the form

$$\phi_{\mathbf{G}}(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{\mathrm{i}\mathbf{G}\cdot\mathbf{r}}. \tag{3.48}$$

Expanding the eigenfunction of the $\mathbf{k}$-dependent Hamiltonian $\hat{H}_\mathbf{k}$ in terms of $\phi_\mathbf{G}(\mathbf{r})$ ensures that $u_{i\mathbf{k}}(\mathbf{r})$ obeys the periodic boundary conditions by construction. Obtaining the eigenfunctions at a certain $\mathbf{k}$-point in the Brillouin zone reduces to finding the lowest $N/2$ eigenstates of a matrix with the elements

$$H_\mathbf{k}^{\mathbf{G},\mathbf{G}'} = \langle \phi_\mathbf{G} | \hat{H}_\mathbf{k} | \phi_{\mathbf{G}'} \rangle \tag{3.49}$$

Plane waves have the convenient property that they form an orthonormal basis that gives an unbiased sampling of the entire space. Furthermore, the action of the kinetic energy operator of the Kohn-Sham Hamiltonian onto a plane wave is trivial to evaluate. The accuracy of the plane wave expansion of the eigenfunctions is controlled by a single parameter $\mathbf{G}_{\max}$, the maximum reciprocal lattice vector $\mathbf{G}$ included in the basis. Given that the plane wave $|\phi_{\mathbf{G}_{\max}}\rangle$ has a kinetic energy associated with it that can simply be written as $|\mathbf{G}_{\max}|^2/2$, the basis set size is often specified via a kinetic energy cutoff. This defines the maximum kinetic energy the Kohn–Sham electrons are allowed to have and thus provides a physically motivated, systematically improvable accuracy of the basis set representation.

While plane-waves form a very effective basis in an infinite solid in periodic boundary conditions, they are not a suitable choice for isolated molecules[10]. In this scenario, the most efficient basis sets come in the form of localised, atom-centered functions, that are generally non-orthogonal. Common choices include Gaussian functions or pseudoatomic orbitals. The non-orthogonality of the basis functions transforms the Kohn–Sham eigen-

[10]This is mainly due to the fact that the treatment of a molecule in periodic boundary conditions requires the explicit inclusion of vacuum into the unit cell to avoid interaction between periodic images of the system. Since plane waves provide a uniform description of the entire unit cell, they are ineffective in this scenario, as they sample the vacuum to the same degree as the molecule.

value equations into generalised eigenvalue equations of the form

$$\mathbf{H}^{\phi}_{\mathrm{KS}}\mathbf{a}_i = \epsilon_i \mathbf{S}^{\phi}\mathbf{a}_i \tag{3.50}$$

where

$$\begin{aligned} \left(H^{\phi}_{\mathrm{KS}}\right)_{\alpha\beta} &= \langle\phi_\alpha|\hat{H}_{\mathrm{KS}}|\phi_\beta\rangle \\ S^{\phi}_{\alpha\beta} &= \langle\phi_\alpha|\phi_\beta\rangle \end{aligned} \tag{3.51}$$

and $\mathbf{a}_i$ forms the coefficient vector of Kohn–Sham state $|\psi^{\mathrm{KS}}_i\rangle$. The localisation of the basis functions allows for efficient calculations on systems of isolated molecules. Furthermore, far fewer basis functions are generally needed to represent the eigenstates of the Hamiltonian compared to the plane-wave basis. However, localised basis sets of the form discussed here do not generally offer the systematic convergence properties of plane-wave basis sets and require special techniques to treat problems arising from non-orthogonality. Furthermore, any localised basis set is naturally unsuited to represent very delocalised Kohn–Sham states. This issue will become important in the discussion of excited states.

3.2.4 Fast-Fourier transforms and Ewald summation

In section 3.2.1 it is pointed out that to obtain the DFT eigenstates of an infinite, periodic system, it is sufficient to solve for the functions $|u_{i\mathbf{k}}\rangle$ in a single primitive real space unit cell for some discrete mesh of points $\mathbf{k}$ confined to the first Brillouin zone in reciprocal space. The periodicity of the Kohn–Sham potential allows for the use of a number of special techniques to speed up calculations. For every density $\rho(\mathbf{r})$ constructed in the self-consistent solution to the Kohn–Sham equations, a new Kohn–Sham potential has to be constructed. For semi-local exchange-correlation functionals, the application of the Hartree potential $V_{\mathrm{H}}(\mathbf{r})$ to a certain Kohn–Sham band is normally the most computationally demanding term to calculate. For a given density, the Hartree potential is simply the convolution of the density with the Coulomb potential in real space

$$V_{\mathrm{H}}(\mathbf{r}) = \int \mathrm{d}^3 r' \, \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.52}$$

The density must have the same periodicity as the lattice and can thus be expanded using a plane wave basis as defined in 3.2.3. The density in real space and the basis set coefficients $\rho(\mathbf{G})$ are related by a simple Fourier transform. However, in reciprocal space,

the convolution in (3.52) reduces to a simple product[11]

$$V_{\mathrm{H}}(\mathbf{G}) = \frac{4\pi\rho(\mathbf{G})}{\Omega|\mathbf{G}|^2} \tag{3.53}$$

where Ω denotes the unit cell volume. Fast-Fourier Transforms (FFTs) [49] allow to change between reciprocal space and real space in a computational effort scaling as $\mathcal{O}(N\log(N))$ with unit cell size, allowing for very efficient calculations of the Hartree potential.

Another term that does need to be evaluated to calculate the total energy per unit cell of the system is the electrostatic energy related to the ion-ion interaction associated with a single unit cell of the crystal.

$$E_{\text{ion-ion}} = \frac{1}{2}\sum_{A}\sum_{B\neq A}\sum_{\mathbf{R}}\frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B + \mathbf{R}|} \tag{3.54}$$

Here, the first two sums run over all the ions in the unit cell, $\mathbf{R}_A$ and $\mathbf{R}_B$ denote the positions of ion A and B and $\mathbf{R}$ denotes a real space lattice vector. The sum over $\mathbf{R}$ runs over an infinite number of real space lattice vectors and accounts for all possible interactions between the primitive unit cell and the infinite number of periodic repeats. Since the Coulomb potential falls off slowly with distance, the sum over $\mathbf{R}$ converges very slowly and is inefficient to evaluate for practical purposes. A more efficient approach comes in the form of the Ewald summation technique [50], where the Coulomb interaction $1/r$ is split into a long range and a short range part such that

$$\frac{1}{r} = \frac{\mathrm{erfc}(\alpha r)}{r} + \frac{\mathrm{erf}(\alpha r)}{r} \tag{3.55}$$

for some Ewald splitting parameter α. This allows the sum over $\mathbf{R}$ to be split into a short range part converging rapidly in real space and a long range part converging rapidly in reciprocal space, yielding a much more efficient computational scaling of $\mathcal{O}(N^2)$ with the number of ions in the unit cell. In principle, the computational effort that can be made to scale as good as $\mathcal{O}(N\log(N))$ with the number of ions in the unit cell if FFTs are used[12].

3.2.5 Iterative eigensolvers

Once the KS potential is constructed, the lowest $N_{\mathrm{occ}}/2$ eigenstates of H_{KS} have to be found in the iterative solution of the Kohn–Sham equations. In case of an isolated sys-

[11] The $|\mathbf{G}|$=0 component is excluded from the calculation as the overall system is taken to be charge neutral and the term is cancelled by analogous terms in the ion-ion and electron-ion interactions.

[12] Given that the Ewald summation has to be only carried out once at the beginning of a calculation, it is often (eg. in the `ONETEP` method) considered to be unnecessary to calculate it in linear-scaling effort since it does not provide a bottleneck for practical calculations.

tem where the KS states are expanded in a relatively small number of localised basis functions $\{\phi_\alpha\}$, it is often possible to directly construct the Hamiltonian matrix $\mathbf{H}^\phi_{\mathrm{KS}}$ and use standard linear-algebra methods to diagonalise it. For localised basis functions, the asymptotic scaling of the construction of the Hamiltonian is linear with system size (provided (semi)-local approximations to the exchange-correlation potential are used), while the diagonalisation techniques generally scale as $\mathcal{O}(N_\phi^3)$ with the number of basis functions N_ϕ.

In an infinite crystal, where plane waves form an appropriate basis set as discussed in section 3.2.3, the computational cost of this direct approach of solving the Kohn–Sham equations is often prohibitive. In a plane wave basis N_ϕ is given by

$$N_\phi \approx \frac{V}{2\pi^2}\frac{|\mathbf{G}_{\max}|^3}{\sqrt{(2^3)}} \tag{3.56}$$

which is generally much larger than the size of localised basis sets. For large cell volumes, it can become prohibitive to construct, store and diagonalise $\hat{H}^{\mathbf{G},\mathbf{G}'}_{\mathbf{k}}$. The explicit construction of $H^{\mathbf{G},\mathbf{G}'}_{\mathbf{k}}$ also scales as $\mathcal{O}(N_\phi^2)$ as the plane waves are delocalised over the entire unit cell. Furthermore, since $N_{\mathrm{occ}}/2 \ll N_\phi$, direct diagonalisation of $H^{\mathbf{G},\mathbf{G}'}_{\mathbf{k}}$ yields a large number of eigenstates that are not needed for the self-consistent solution of the KS equations.

The above problems can be avoided by using an iterative eigensolver. Since solving for a subset of eigenvalues of large matrices is a standard problem in computational science, a variety of different algorithms exist that vary in convergence properties and memory requirements [51]. Among the most popular in electronic structure theory are the Davidson [52] and Conjugate Gradients algorithms [53], which are based on minimising the Rayleigh quotient

$$\rho(\mathbf{x}) = \sum_i^{N_{\mathrm{occ}}/2} \frac{\mathbf{x}_i^\dagger \mathbf{H}^\phi_{\mathrm{KS}} \mathbf{x}_i}{\mathbf{x}_i^\dagger \mathbf{S}^\phi \mathbf{x}_i} \tag{3.57}$$

under the constraint

$$\mathbf{x}_i^\dagger \mathbf{S}^\phi \mathbf{x}_j = \delta_{ij} \tag{3.58}$$

where $\mathbf{H}^\phi_{\mathrm{KS}}$ is again the Hamiltonian in some basis set representation ϕ, $\{\mathbf{x}_i\}$ is a set of approximate eigenvectors and $\mathbf{S}^\phi$ is the overlap matrix of the basis set. Minimising (3.57) with respect to $\{\mathbf{x}_i\}$ yields the lowest $N_{\mathrm{occ}}/2$ eigenstates of $\mathbf{H}^\phi_{\mathrm{KS}}$. A main feature of iterative methods is that to perform the minimisation it is sufficient to know the action of the Hamiltonian on an arbitrary trial vector $\mathbf{x}$. Thus in practice, one never has to explicitly calculate or store $\mathbf{H}^\phi_{\mathrm{KS}}$, which is prohibitive for large basis set representations. For localised basis sets, $\mathbf{H}^\phi_{\mathrm{KS}}$ is sparse and computing $\mathbf{H}^\phi_{\mathrm{KS}}\mathbf{x}_i$ can be made to scale linearly

with the number of basis functions. However, keeping $\{\mathbf{x}_i\}$ orthogonal has a computational complexity of $\mathcal{O}(2N_\phi(N_{\text{occ}}/2)^2)$ associated with it[13], leading to a still overall cubic scaling with system size when solving the Kohn–Sham equations.

3.2.6 Norm-conserving pseudopotentials

Close to the ionic core, the Coulomb potential diverges and the Kohn–Sham valence wavefunctions are known to show rapid oscillations in order to fulfil the orthogonality requirement to the core wavefunctions. This causes numerical problems in representing these states, especially when plane wave basis sets are used. To resolve the rapid oscillations close to the ionic core, very high kinetic energy cutoffs are required, making basis sets prohibitively large for practical calculations. To address this issue, plane-wave DFT codes generally make use of pseudopotentials.

In the pseudopotential method, the chemically inert core electrons are removed from the calculation and the bare ionic potential is replaced by a soft pseudopotential that reproduces the effect of the core electrons on the valence wave functions outside a certain cutoff region, while yielding smooth valence wave functions inside the core region. Since the core electrons to a good approximation do not take part in chemical bonding, a well constructed pseudopotential produces the right behaviour of Kohn-Sham valence wavefunctions outside the core regions no matter what chemical environment the ion is placed in.

Pseudopotentials are generally generated by solving a radial Kohn-Sham equation of the form

$$\left[-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}r}+\frac{l(l+1)}{r^2}+V_{\text{KS}}[\rho](r)\right]rR_{nl}(r)=\epsilon_{nl}rR_{nl}(r) \tag{3.59}$$

for each atomic species in isolation. Here, l determines the angular momentum quantum number, $R_{nl}(r)$ are the radial wavefunctions and $\rho(r)$ is the radial density associated with the sum of the occupied wavefunctions $R_{nl}(r)$. For the purpose of this work, only norm-conserving pseudopotentials are considered. These so called norm-conserving pseudopotentials have a number of desirable features which can be summarised as [54]:

- The valence pseudo-wavefunctions generated from the pseudopotentials should contain no nodes. This is to ensure computational efficiency, as nodes are generally associated with high kinetic energy components that are to be avoided to keep the required plane-wave basis set as small as possible.

- The radial all-electron wavefunction of a valence state n is equal to the radial pseudo

[13]The computational complexity corresponds to that of the Gram-Schmidt method, a commonly used technique to orthonormalise a set of vectors.

wavefunction outside a given core cutoff region defined by r_c:

$$R_{nl}^{\mathrm{AE}}(r) = R_{nl}^{\mathrm{PS}}(r) \text{ for } r > r_c \tag{3.60}$$

- The charge enclosed in the core region must be equal for both the all-electron and the pseudo wavefunction such that:

$$\int_0^{r_c} |R_{nl}^{\mathrm{AE}}(r)|^2 r^2 \, \mathrm{d}r = \int_0^{r_c} |R_{nl}^{\mathrm{PS}}(r)|^2 r^2 \, \mathrm{d}r \tag{3.61}$$

  This is the definition of "norm-conserving" when referring to pseudopotentials.

- Both the pseudo Kohn-Sham wavefunction and the all-electron wavefunction have the same eigenvalue associated with them

$$\epsilon_{nl}^{\mathrm{AE}} = \epsilon_{nl}^{\mathrm{PS}} \tag{3.62}$$

In general, there exists a tradeoff between the "softness" and the transferability of a generated normconserving pseudopotential. While large core region cutoffs r_c lead to very soft pseudopotentials that produce valence Kohn-Sham wavefunctions representable by only few plane waves, harder pseudopotentials with small r_c are much more reliable when being transferred to different chemical environments.

3.2.7 Projector augmented-wave formalism

While the norm-conserving pseudopotential approach is relatively simple and intuitive, it comes with a number of drawbacks. First, many transition metals require the explicit treatment of semi-core states as valence states in order to achieve satisfactory agreements with experimental data, resulting in very hard pseudopotentials. Second, any access to the all-electron density and the full all-electron wavefunctions is lost in the pseudopotential approach. The projector augmented-wave (PAW) formalism [55] has been developed to address these issues by defining a linear transformation τ that maps soft pseudo (PS) Kohn-Sham wavefunctions onto their all-electron (AE) equivalents such that

$$|\psi\rangle = \tau|\tilde{\psi}\rangle \tag{3.63}$$

Here, quantities with a tilde denote quantities associated with the PS Hilbert space. Given that the all-electron wavefunctions outside the core region are smooth quantities, we require the AE wavefunctions and the pseudo-wavefunctions to differ only inside the core regions of the ions. Inside these augmentation regions, the operator is expanded

by defining AE partial waves $\{|\varphi_i\rangle\}$ that are linked to PS partial waves $\{|\tilde{\varphi}_i\rangle\}$ by the operator τ inside the augmentation region. It is required that $|\varphi_i\rangle = |\tilde{\varphi}_i\rangle$ outside the augmentation region and that $\{|\tilde{\varphi}_i\rangle\}$ is a set of numerically convenient smooth functions. The partial waves are generally written in terms of a radial function expressible on a radial grid around the atoms within the augmentation region and a spherical harmonic, such that

$$\varphi_j(\mathbf{r}) = \frac{\varphi_{n_j,L_j}(r)}{r} S_{L_j,M_j}(\hat{\mathbf{r}}). \tag{3.64}$$

The linear transformation τ can then be written as

$$\tau = 1 + \sum_i \left(|\varphi_i\rangle - |\tilde{\varphi}_i\rangle\right) \langle \tilde{p}_i| \tag{3.65}$$

where $\langle \tilde{p}_i|$ are fixed projector functions for each partial wave. The projector functions are required to fulfil the completeness and orthonormality conditions

$$\sum_i |\tilde{\varphi}_i\rangle\langle \tilde{p}_i| = 1 \tag{3.66}$$

$$\langle \tilde{p}_i|\tilde{\varphi}_j\rangle = \delta_{ij}. \tag{3.67}$$

The full mapping between the AE wavefunction and the PS wavefunction can thus be expressed as

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_i \left(|\varphi_i\rangle - |\tilde{\varphi}_i\rangle\right) \langle \tilde{p}_i|\tilde{\psi}\rangle. \tag{3.68}$$

As required, the AE wavefunction reduces to the PS wavefunction outside the augmentation spheres around the ions, while inside the spheres the soft part is replaced by an all-electron part defined via the projectors $\langle \tilde{p}_i|$ and the AE partial waves $|\varphi_i\rangle$.

Using the linear operator τ the Kohn–Sham eigenvalues of the all-electron Hamiltonian can be expressed as

$$\frac{\langle \tilde{\psi}_i|\tau^\dagger \hat{H}^{\mathrm{KS}} \tau|\tilde{\psi}_i\rangle}{\langle \tilde{\psi}_i|\tau^\dagger \tau|\tilde{\psi}_i\rangle} = \epsilon_i^{\mathrm{KS}}. \tag{3.69}$$

Therefore, in order to obtain the all-electron Kohn–Sham eigenvalues, it is sufficient to solve the Kohn–Sham equations for the smooth pseudo-wavefunctions and an effective transformed Hamiltonian $\tilde{H} = \tau^\dagger \hat{H}_{\mathrm{KS}} \tau$, where the pseudo-wavefunctions need to obey a modified orthonormality condition $\langle \tilde{\psi}_i|\tau^\dagger \tau|\tilde{\psi}_j\rangle = \delta_{ij}$. The power of the formalism lies in the fact that all-electron quantities like the total all-electron energy and bandstructure are directly accessible while one only has to solve for the smooth pseudo Kohn–Sham eigenstates which are conveniently expressed by plane wave functions.

3.3 Linear-scaling density-functional theory

As seen in section 3.1.3, in order to achieve an accurate representation of the kinetic energy functional, one has to introduce $N_{\mathrm{occ}}/2$ orthogonal eigenstates $|\psi_i^{\mathrm{KS}}\rangle$ of a fictitious non-interacting system. Solving the Kohn–Sham equations self-consistently for these eigenstates introduces an asymptotic scaling of $\mathcal{O}(N^3)$ with the number of electrons due to the orthonormality constraint. This unfavourable scaling places severe limitations on the sizes of systems that can be studied using conventional Kohn–Sham DFT. For this reason there has been a long-lasting research effort into methods that show a *linear* asymptotic scaling with system size [56, 57], while retaining the accuracy of orbital-based DFT[14]. In the following, some of the main concepts of a particular class of linear-scaling DFT methods, namely methods based on the single-particle density matrix, are introduced. The discussion presented here is limited to systems with a finite band gap.

3.3.1 The single-particle density matrix

The most successful subset of linear-scaling DFT methods over the recent years relies on the use of the single-particle density matrix. Since the orthogonality constraint on the Kohn–Sham orbitals poses a major problem for linear-scaling methods, one has to give up the use of individual Kohn-Sham states in favour of a collective representation. The single-particle Kohn–Sham density matrix can be introduced as

$$\rho^{\{v\}}(\mathbf{r},\mathbf{r}') = \sum_i^{N_{\mathrm{occ}}/2} \psi_i^{\mathrm{KS}*}(\mathbf{r})\psi_i^{\mathrm{KS}}(\mathbf{r}') \tag{3.70}$$

such that the KS density of the system can be written as $\rho(\mathbf{r}) = 2\rho^{\{v\}}(\mathbf{r},\mathbf{r})$. In order to be a valid density matrix of an independent particle Kohn–Sham system, $\rho^{\{v\}}(\mathbf{r},\mathbf{r}')$ has to obey two conditions that directly follow from equation (3.70):

$$\int \mathrm{d}^3r''\, \rho^{\{v\}}(\mathbf{r},\mathbf{r}'')\rho^{\{v\}}(\mathbf{r}'',\mathbf{r}') = \rho^{\{v\}}(\mathbf{r},\mathbf{r}') \tag{3.71}$$

$$2\int \mathrm{d}^3r\, \rho^{\{v\}}(\mathbf{r},\mathbf{r}) = N_{\mathrm{occ}}. \tag{3.72}$$

The first condition is known as the idempotency condition [58] and follows directly from the orthogonality of the Kohn–Sham states. The second condition is a normalisation constraint that ensures that the density associated with the single-particle density matrix integrates to the number of (valence)-electrons in the system. Therefore, in solving for the

[14] As mentioned earlier, OF-DFT trivially obtains an $\mathcal{O}(N)$ scaling with system size but, due to the approximation of the kinetic energy functional, lacks the accuracy of Kohn–Sham approaches.

density matrix rather than the individual KS eigenstates, the orthonormality constraint on the eigenstates can be dropped and replaced by the idempotency and normalisation constraints of the density matrix.

3.3.2 Localised representation and non-orthogonality

While the collective representation of KS states in terms of a density matrix avoids the orthogonality constraint on individual eigenstates in favour of a single idempotency constraint, this approach can only lead to efficient algorithms if the density matrix can be represented in an efficient basis. In linear-scaling approaches, the density matrix is generally expanded in a set of localised, non-orthogonal orbitals or support functions $\{\phi_\alpha\}$. One then writes

$$\rho^{\{v\}}(\mathbf{r},\mathbf{r}') = \sum_{\alpha\beta} \phi^*_\alpha(\mathbf{r}) P^{\{v\}\alpha\beta} \phi_\beta(\mathbf{r}') \tag{3.73}$$

where $\mathbf{P}^{\{v\}}$ is the ground state KS density matrix in the representation of localised orbitals. While for isolated molecular systems, it is always possible to choose a localised representation in the form of equation (3.73), the validity of the expression for infinite systems in periodic boundary conditions crucially depends on the existence of exponentially localised Wannier functions in those systems. As established in section 3.2.2, in an insulating system the valence Kohn–Sham states form a composite band for which a set of exponentially localised Wannier functions can be constructed. From this it directly follows that a representation of the form of equation (3.73) is generally possible even in periodic systems, as long as the system in question is insulating. Note that while the formal justification for the localised representation in (3.73) originates from the existence of Wannier functions, the support functions $\{\phi_\alpha\}$ are generally not chosen to be Wannier functions constructed as described in section 3.2.2, since that construction would require the self-consistent solution of the KS equations for all KS eigenstates. In most methods, more generalised localised functions like atom-centered pseudoatomic orbitals are chosen [59, 60] in order to represent the valence density matrix. This results in the fact that the set of support functions chosen in order to obtain a good representation of $\rho^{\{v\}}(\mathbf{r},\mathbf{r}')$ is generally larger than the minimal set of Wannier functions necessary to span the occupied subspace.

Since the localised support functions are generally non-orthognonal, it is convenient to introduce a differentiation between covariant and contravariant tensorial quantities. Quantities with subscript greek letters are taken to be to be covariant quantities, while superscript greek letters are contravariant quantities. The covariant and contravariant support functions form a dual set such that $\langle\phi_\alpha|\phi^\beta\rangle = \delta^\beta_\alpha$. The overlap matrix $S^\phi_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$ is an example of a covariant quantity, while $\mathbf{P}^{\{v\}}$ and the inverse of the overlap

matrix $(\mathbf{S}^\phi)^{-1}$ are contravariant quantities. In this sense Equation (3.73) is an example of a contraction between contravariant and covariant tensor quantities. In the following sections, a repeated greek index on a contra- and covariant quantity will be taken to mean a contraction of the form (3.73) and any explicit sums over localised orbitals are dropped (for more details on the tensorial nature of non-orthogonal basis sets in electronic structure theory, see [61, 62]).

3.3.3 Density matrix DFT

The self-consistent solution of the Kohn–Sham equations can be replaced by a single energy minimisation in the density matrix formalism. Choosing the localised support functions as a representation for the single-particle Kohn–Sham Hamiltonian, the band-structure energy $E_{\mathrm{BS}}[\rho]$ can be written as

$$E_{\mathrm{BS}}[\rho] = 2\mathrm{Tr}\left[\mathbf{P}^{\{v\}}\mathbf{H}^{\phi}_{\mathrm{KS}}\right] \tag{3.74}$$

Therefore, by using the variational principle, carrying out a direct minimisation of the bandstructure energy with respect to $\mathbf{P}^{\{v\}}$

$$E_{\mathrm{BS}}[\rho_0] = \min_{\mathbf{P}^{\{v\}}} 2\mathrm{Tr}\left[\mathbf{P}^{\{v\}}\mathbf{H}^{\phi}_{\mathrm{KS}}\right] \tag{3.75}$$

subject to the constraints

$$\left(P^{\{v\}}S^{\phi}P^{\{v\}}\right)^{\alpha\beta} = P^{\{v\}\alpha\beta} \tag{3.76}$$

$$2\mathrm{Tr}\left[\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\right] = N_{\mathrm{occ}} \tag{3.77}$$

yields the Kohn–Sham ground state density ρ_0 and the ground state energy of the system without any reference to individual Kohn–Sham states and their orthogonality. Note however that $\mathbf{H}^{\phi}_{\mathrm{KS}}$ is dependent on the Kohn–Sham density and thus on the density matrix. Thus the minimisation has to be performed self-consistently, just like in standard Kohn–Sham approaches. Self-consistency can be achieved by minimizing $E_{\mathrm{BS}}[\rho_0]$ for a fixed $\mathbf{H}^{\phi}_{\mathrm{KS}}$. The resulting idempotent, normalised $\mathbf{P}^{\{v\}}$ generates a new ρ and thus a new $\mathbf{H}^{\phi}_{\mathrm{KS}}$ and the optimisation can be repeated until self-consistency is reached.

The normalisation constraint to the density matrix can be incorporated into the minimisation scheme by considering the grand potential Ω rather than the band structure energy [63]

$$\Omega[\mathbf{P}^{\{v\}}] = 2\mathrm{Tr}\left[\mathbf{P}^{\{v\}}\left(\mathbf{H}^{\phi}_{\mathrm{KS}} - \mu\mathbf{S}^{\phi}\right)\right] \tag{3.78}$$

and minimising it at fixed chemical potential μ. However, carrying out the minimisation

while simultaneously enforcing the idempotency constraint on the ground state density matrix requires special numerical techniques and algorithms to ensure the stability of the minimisation procedure, which is discussed in the next sections.

3.3.4 Purification and penalty functionals

As was pointed out by McWeeny [58], it is possible to drive a near idempotent density matrix to idempotency using a purification scheme. Consider the functional $Q[\mathbf{P}^{\{v\}}]$ defined as the square of the deviation from idempotency of $\mathbf{P}^{\{v\}}$:

$$Q[\mathbf{P}^{\{v\}}] = \mathrm{Tr}\left[\left(\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi} - \mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\right)^{2}\right] \tag{3.79}$$

The functional is necessarily positive-semidefinite and strictly vanishes for idempotent $\mathbf{P}^{\{v\}}$. Differentiating the scalar quantity $Q[\mathbf{P}^{\{v\}}]$ with respect to $\mathbf{P}^{\{v\}}$ one obtains the covariant gradient:

$$\begin{aligned}\frac{\partial Q[\mathbf{P}^{\{v\}}]}{\partial \mathbf{P}^{\{v\}}} &= 4\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi} \\ &\quad -6\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi} + 2\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\end{aligned} \tag{3.80}$$

The covariant gradient can be changed into a contravariant gradient with the same tensorial properties as $\mathbf{P}^{\{v\}}$ by multiplying the above expression from the left and right with the inverse overlap matrix. Using the contravariant gradient as a steepest descent direction for improving a trial guess $\mathbf{P}_{n}^{\{v\}}$, $\mathbf{P}_{n+1}^{\{v\}}$ can be written as

$$\mathbf{P}_{n+1}^{\{v\}} = 3\mathbf{P}_{n}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}_{n}^{\{v\}} - 2\mathbf{P}_{n}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}_{n}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}_{n}^{\{v\}} \tag{3.81}$$

It is then in principle possible to find the DFT ground state energy by performing a two-step approach of alternatingly minimising $\Omega[\mathbf{P}^{\{v\}}]$ and $Q[\mathbf{P}^{\{v\}}]$. However, this approach is potentially numerically unstable and an alternative method comes in the form of including the effect of $Q[\mathbf{P}^{\{v\}}]$ into the ground state energy minimisation via a penalty functional. Consider the effective total energy functional introduced by Kohn [64]

$$F_{1}[\mathbf{P}^{\{v\}}] = \Omega[\mathbf{P}^{\{v\}}] + \alpha\left[Q[\mathbf{P}^{\{v\}}]\right]^{\frac{1}{2}} \tag{3.82}$$

where α is some positive constant. The functional $F_{1}[\mathbf{P}^{\{v\}}]$ has a true minimum for the idempotent ground state density matrix. This follows from the fact that the energy is non-variational with respect to changes in the occupancies of $\mathbf{P}^{\{v\}}$ due to the presence of the square root. The square root causes the functional to have a branch point at

the correct occupancy, which causes issues for standard conjugate gradient minimisation schemes [65, 66].

The alternative to the Kohn functional is to consider [66]

$$F_2[\mathbf{P}^{\{v\}}] = \Omega[\mathbf{P}^{\{v\}}] + \alpha Q[\mathbf{P}^{\{v\}}]. \tag{3.83}$$

This functional does not contain any branch points and is easy to minimise with standard iterative techniques. However, the missing branch point causes F_2 to only have a minimum at an approximately idempotent density matrix $\mathbf{P}'^{\{v\}}$, with $\mathbf{P}'^{\{v\}}$ approaching the correct idempotent ground state density matrix as $\alpha \to \infty$. The error in the occupancies of $\mathbf{P}'^{\{v\}}$ can be shown to be proportional to α^{-1}, causing significant errors in the bandstructure energy for practical values of α. However, it is possible to derive a correction term to the energy due to the imperfect idempotency such that [66]

$$E_{\mathrm{BS}}[\rho_0] \approx E_{\mathrm{BS}}[\mathbf{P}'^{\{v\}}] - 2\alpha \mathrm{Tr}\left[\mathbf{P}'^{\{v\}}\mathbf{S}^{\phi}\left(1 - \mathbf{P}'^{\{v\}}\mathbf{S}^{\phi}\right)^2\left(1 - 2\mathbf{P}'^{\{v\}}\mathbf{S}^{\phi}\right)\right]. \tag{3.84}$$

3.3.5 The LNV method

While enforcing the density matrix idempotency with penalty functionals causes a number of problems in standard minimisation approaches, it was shown by Li, Nunes and Vanderbilt [63] that these issues can be largely avoided. In the LNV method, one introduces a trial or auxiliary density matrix $\mathbf{L}^{\{v\}}$ such that

$$\mathbf{P}^{\{v\}} = 3\mathbf{L}^{\{v\}}\mathbf{S}^{\phi}\mathbf{L}^{\{v\}} - 2\mathbf{L}^{\{v\}}\mathbf{S}^{\phi}\mathbf{L}^{\{v\}}\mathbf{S}^{\phi}\mathbf{L}^{\{v\}} \tag{3.85}$$

ie. $\mathbf{P}^{\{v\}}$ is the first order idempotency corrected version of $\mathbf{L}^{\{v\}}$ following equation (3.81). Now, one considers the grand potential Ω to be a function of $\mathbf{L}^{\{v\}}$ such that

$$\Omega[\mathbf{L}^{\{v\}}] = \mathrm{Tr}\left[\left(3\mathbf{L}^{\{v\}}\mathbf{S}^{\phi}\mathbf{L}^{\{v\}} - 2\mathbf{L}^{\{v\}}\mathbf{S}^{\phi}\mathbf{L}^{\{v\}}\mathbf{S}^{\phi}\mathbf{L}^{\{v\}}\right)\left(\mathbf{H}^{\phi}_{\mathrm{KS}} - \mu\mathbf{S}^{\phi}\right)\right]. \tag{3.86}$$

The functional $\Omega[\mathbf{L}^{\{v\}}]$ has a *local* minimum where $\mathbf{P}^{\{v\}}$ corresponds to the idempotent Kohn–Sham ground state density of the system. The global minimum is characterised by a runaway solution where the occupancies of states below and above the Fermi level approach positive and negative infinity respectively. However, as long as a near idempotent density matrix is used as an input guess, with occupancies close to 0 and 1, standard iterative minimisation techniques applied to equation (3.86) yield the idempotent ground state density matrix corresponding to the Hamiltonian $\mathbf{H}^{\phi}_{\mathrm{KS}}$.

3.3.6 Linear-scaling DFT

In the previous sections it was shown that density matrix DFT avoids the source $\mathcal{O}(N^3)$ scaling of standard Kohn–Sham DFT by replacing the orthonormality of Kohn–Sham states with the idempotency of the single-particle density matrix. However, regardless of whether idempotency is achieved via a penalty functional minimisation or the LNV method, it generally involves the multiplication of density matrices. Since $\{\phi_\alpha\}$ and thus the dimensions of $\mathbf{L}^{\{v\}}$ or $\mathbf{P}^{\{v\}}$ grow linearly with system size, the computational effort in computing matrix products of the form $\mathbf{L}^{\{v\}}\mathbf{S}^{\phi}\mathbf{L}^{\{v\}}$ generally scales as $\mathcal{O}(N^3)$ with system size, suggesting that the density matrix formalism does not in fact yield any improvement in scaling over the conventional KS-DFT method.

However, the considerations that prove the existence of maximally localised Wannier functions for a set of composite bands (see section 3.2.2) also show that the single particle density matrix corresponding to the same set of composite bands in a crystal with inversion symmetry decays exponentially with distance [46], such that

$$\lim_{|\mathbf{r}-\mathbf{r}'|\to\infty} e^{\epsilon A|\mathbf{r}-\mathbf{r}'|}\rho^{\{v\}}(\mathbf{r},\mathbf{r}') = 0 \tag{3.87}$$

for some $0 < \epsilon < 1$. Thus in an insulator, the Kohn–Sham single particle density matrix decays exponentially with distance and the rate of the exponential decay A is a function of the band gap [67]. Since the density matrix is represented by localised functions, it follows that in the case of insulators $\mathbf{L}^{\{v\}}$ and $\mathbf{P}^{\{v\}}$ are sparse for sufficiently large system sizes. In the limit of large enough systems, the resulting matrices of all matrix products required in the energy minimisation become sparse and all matrix operations can be performed in $\mathcal{O}(N)$ effort.

Linear-scaling DFT calculations are thus possible if the explicit calculation of individual Kohn–Sham states is abandoned in favour of a collective representation through the single particle density matrix and if the sparsity of the density matrix is exploited. The nature of the density matrix representation means that some quantities readily accessible in Kohn–Sham DFT like the Kohn-Sham bandstructure of the system are inaccessible in the linear-scaling formulation, although they can generally be obtained in a single $\mathcal{O}(N^3)$ post-processing step of diagonalising the Kohn–Sham Hamiltonian.

Chapter 4

Approximations to excited states

In this chapter, the focus lies on the approximate evaluation of excited states in the many-electron system. The discussion is limited to neutral excitations in which the many-electron system couples to light by absorbing photons of specific wavelengths. In a quasiparticle or Kohn–Sham picture, this photon absorption event can be generally described through the creation of an electron-hole pair in the system. In this chapter, two main approaches to this problem are compared and contrasted, one being the time-dependent extension to DFT (TDDFT) and the other being the Green's function based approach of many-body perturbation theory (MBPT). The chapter concludes with some scaling considerations that assess the appropriateness of either approach for the large-scale applications of interest in this work. As in the previous chapter, the only systems considered are those with a spin-degenerate ground state. All spin indices are dropped for convenience and are only reintroduced where the electron spin has to be considered explicitly.

4.1 Time-dependent density-functional theory

As pointed out in section 3.1, the Hohenberg–Kohn theorem establishes a one-to-one mapping between the ground state density of a many-electron system and the external potential. Since the external potential uniquely defines the many-electron Hamiltonian of a system, in principle any excited state of the full many-body Hamiltonian must itself be a functional of the ground state density. However, the Hohenberg–Kohn theorem does not say anything about the form of such a functional for an arbitrary excited state and therefore does not provide a prescription for obtaining it.

For this reason, DFT cannot be used to yield neutral excitations[15]. While the Kohn–

[15]Here, we limit the discussion to vertical excitations of closed-shell systems. DFT can be used to yield excitation energies in cases where different spin configurations yield different ground state energies. The energy difference between two different configurations can then be interpreted as the excitation energy of the system going from one symmetry to the other.

Sham formalism provides the individual eigenstates of a fictitious non-interacting system, the Kohn–Sham energy differences cannot be treated as neutral excitation energies, even if the Kohn–Sham energies are interpreted as quasiparticle bandstructure energies. One could consider the lowest excitation of a system to consist of an electron being moved from the highest occupied Kohn–Sham state to the lowest unoccupied state and their energy difference to be the excitation energy. However, this interpretation is not only lacking any formal justification, it is also wrong, as in this system the external potential is fixed and thus not appropriate for such an excitation.

One way of introducing the possibility of excitations into the Kohn–Sham formalism is to explicitly consider the Hamiltonian of a system coupling to an external electromagnetic field. One can then consider the general form of the non-relativistic time-dependent Hamiltonian

$$\hat{H}_{\mathrm{EM}}(t) = \frac{1}{2}\sum_i \left[\nabla_i + \mathbf{A}_{\mathrm{ext}}(\mathbf{r}_i, t)\right]^2 + \sum_i V_{\mathrm{ext}}(\mathbf{r}_i, t) + \hat{V}_{\mathrm{ee}} \tag{4.1}$$

where $V_{\mathrm{ext}}(\mathbf{r}_i, t)$ and $\mathbf{A}_{\mathrm{ext}}(\mathbf{r}_i, t)$ are the external scalar and vector potentials experienced by particle i at time t and $\hat{V}_{\mathrm{ee}}$ again denotes the electron-electron interaction. The above form of the Hamiltonian is capable of describing interactions between the many-electron system and an external, *classical* electromagnetic field and is therefore the most general form to describe non-relativistic light-matter interactions. Note that the Hamiltonian is now explicitly time-dependent and thus cannot be treated with the techniques established through the Hohenberg–Kohn theorem.

A major complication in the formal treatment of (4.1) comes through the introduction of the vector potential $\mathbf{A}_{\mathrm{ext}}$. Thus in practice, further approximations are introduced to yield a simpler effective Hamiltonian. First, note that the original Hamiltonian in equation (4.1), like the Hamiltonian considered for the ground state problem in the previous section, does not contain any explicit spin-dependence and thus spin-orbit coupling effects as well as the explicit coupling of the electron spin to the external magnetic field are ignored. This means that a range of excitations often encountered in magnetic materials, like collective spin-wave excitations, are ignored in the Hamiltonian of equation (4.1). Furthermore, any fine splitting of excitations due to spin-orbit coupling effects can naturally not be obtained from the spin-independent Hamiltonian[16]. Now, in a further approximation to equation (4.1) one can ignore all interaction effects between the current of moving electrons and the external magnetic field, thus removing any coupling to magnetic effects from the system as well as the need to treat the vector potential $\mathbf{A}_{\mathrm{ext}}$.

[16] Although, if spin-orbit effects are small, which is the requirement for the Hamiltonian in equation (4.1) to be appropriate for the system at hand, the fine splitting can be obtained perturbatively (see for example [68]).

As a next step, consider that for low energy excitations, the wavelength of an electromagnetic wave is generally much larger than the spatial extent of the system of interest, so that one can treat the applied electric field experienced by the electrons of the system as constant in space. This approximation is known as the dipole approximation. One can then write the perturbing potential to the ground state system felt by electron i due to an externally applied time-varying electric field $\mathbf{E_{ext}}(t)$ as

$$\delta V_{\text{pert}}(\mathbf{r}_i, t) = -\mathbf{E_{ext}}(t) \cdot \hat{\mathbf{r}}_i \tag{4.2}$$

and thus consider a time-dependent external field

$$V_{\text{ext}}(\mathbf{r}_i, t) = -\sum_A \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \delta V_{\text{pert}}(\mathbf{r}_i, t) = V_{\text{ext}}(\mathbf{r}_i) + \delta V_{\text{pert}}(\mathbf{r}_i, t) \tag{4.3}$$

where $V_{\text{ext}}(\mathbf{r}_i)$ is the usual static external potential considered in the ground state DFT formalism.

Under the approximations mentioned above, it is therefore sufficient to consider a much simpler Hamiltonian:

$$\hat{H}(t) = \hat{T} + \hat{V}_{ee} + \sum_i V_{\text{ext}}(\mathbf{r}_i, t). \tag{4.4}$$

Note that this simplified Hamiltonian is identical to the one addressed in the original Hohenberg–Kohn theorem, with the difference that the external scalar potential is now explicitly dependent on time, due to the introduction of the time-dependent perturbing potential $\delta V_{\text{pert}}(\mathbf{r}_i, t)$. A formally justified extension of the Hohenberg–Kohn theorem to time-varying external scalar potentials is known as the Runge–Gross theorem [69], leading to time-dependent density-functional theory (TDDFT).

It should be pointed out that the original Hamiltonian defined in equation (4.1) can also be treated in a Hohenberg–Kohn type framework, yielding time-dependent current density functional theory (TD-CDFT) [70, 71]. However, since for the purpose of this work, all effects due to a coupling of the electrons to magnetic fields are ignored, the treatment of time-dependent perturbations of the form of δV_{ext} is sufficient.

4.1.1 The Runge–Gross theorem

As an extension to the Hohenberg–Kohn theorem established in 3.1, in their influential paper [69] Runge and Gross provide a proof for the following statements:

- There exists a one-to-one mapping between the external time-dependent potential $V_{\text{ext}}(\mathbf{r}, t)$ and the electron density $\rho(\mathbf{r}, t)$ (up to a scalar function $C(t)$) for an inter-

acting many-particle system evolving from a fixed initial state $\Psi(t_0) = \Psi_0$, given that $V_{\text{ext}}(\mathbf{r}, t)$ can be expanded as a Taylor series around time t_0.

- The quantum mechanical action integral $A[\Psi]$ can be written as a functional of the density. Furthermore, for a time-dependent external potential V_{ext}, $A[\rho]$ can be written as $A[\rho] = B[\rho] - \int_{t_0}^{t_1} \mathrm{d}t \int \mathrm{d}^3 r \rho(\mathbf{r}, t) V_{\text{ext}}(\mathbf{r}, t)$, where $B[\rho]$ is a universal (ie. system-independent) functional of the density.

To prove the first statement, consider two time-dependent external potentials $V_{\text{ext}}(\mathbf{r}, t)$ and $V'_{\text{ext}}(\mathbf{r}, t)$ such that the two potentials differ by more than a time-dependent function $C(t)$, ie. $V_{\text{ext}}(\mathbf{r}, t) - V'_{\text{ext}}(\mathbf{r}, t) \neq C(t)$. It is assumed that both potentials can be expanded as a Taylor series in time around $t = t_0$. From this it follows that there must be some n-th term of the Taylor series where the Taylor epxansion terms of $V_{\text{ext}}(\mathbf{r}, t)$ and $V'_{\text{ext}}(\mathbf{r}, t)$ differ by more than a constant c. It is therefore implied that [69]

$$\frac{\partial^n}{\partial t^n} \left[V_{\text{ext}}(\mathbf{r}, t) - V'_{\text{ext}}(\mathbf{r}, t) \right]_{t=t_0} \neq c \tag{4.5}$$

for some $n \geq 0$.

It is now useful to introduce the concept of a current density $\mathbf{j}(\mathbf{r}$,t) of the many-electron wavefunction $\Psi(t)$ where one can write

$$\begin{aligned} \mathbf{j}(\mathbf{r}, t) &= \left\langle \Psi(t) \left| \hat{\mathbf{j}}(\mathbf{r}) \right| \Psi(t) \right\rangle \\ &= \frac{1}{2\mathrm{i}} \left[(\nabla_{\mathbf{r}} - \nabla_{\mathbf{r}'}) \int \prod_{j=2}^{N} \mathrm{d}^3 r_j \, \Psi^*(\mathbf{r}, \mathbf{r}_2, \cdots \mathbf{r}_N, t) \Psi(\mathbf{r}', \mathbf{r}_2, \cdots \mathbf{r}_N, t) \right]_{\mathbf{r}=\mathbf{r}'} \end{aligned} \tag{4.6}$$

where $\hat{\mathbf{j}}(\mathbf{r})$ is the current density operator. The time evolution of the current density can then be expressed in terms of an equation of motion

$$\mathrm{i} \frac{\partial \mathbf{j}(\mathbf{r}, t)}{\partial t} = \left\langle \Psi(t) \left| \left[\hat{\mathbf{j}}(\mathbf{r}), \hat{H}(t) \right] \right| \Psi(t) \right\rangle . \tag{4.7}$$

Considering the difference between current densities corresponding to systems specified by the two external potentials $V_{\text{ext}}(\mathbf{r}, t)$ and $V'_{\text{ext}}(\mathbf{r}, t)$ evolving from the same initial state Ψ_0 and substituting into equation (4.7), it follows that at $t = t_0$,

$$\begin{aligned} \mathrm{i} \left. \frac{\partial \left[\mathbf{j}(\mathbf{r}, t) - \mathbf{j}'(\mathbf{r}, t) \right]}{\partial t} \right|_{t=t_0} &= \left\langle \Psi_0 \left| \left[\hat{\mathbf{j}}(\mathbf{r}), \left(\hat{H}(t_0) - \hat{H}'(t_0) \right) \right] \right| \Psi_0 \right\rangle \\ &= \mathrm{i} \rho(\mathbf{r}, t_0) \nabla \left[V_{\text{ext}}(\mathbf{r}, t_0) - V'_{\text{ext}}(\mathbf{r}, t_0) \right] . \end{aligned} \tag{4.8}$$

In last line the fact was used that $\hat{H}(t)$ and $\hat{H}'(t)$ only differ in their external potentials. It is now possible to apply the time evolution equation n times, from which it follows

that [69]

$$\mathrm{i}^{n+1}\left.\frac{\partial^{n+1}\left[\mathbf{j}(\mathbf{r},t)-\mathbf{j}'(\mathbf{r},t)\right]}{\partial t^{n+1}}\right|_{t=t_0}=\rho(\mathbf{r},t_0)\nabla\left[\mathrm{i}^{n+1}\left.\frac{\partial^{n}\left[V_{\mathrm{ext}}(\mathbf{r},t)-V'_{\mathrm{ext}}(\mathbf{r},t)\right]}{\partial t^{n}}\right|_{t=t_0}\right] \tag{4.9}$$

It was however established in equation (4.5) that there exists an n-th term in the Taylor expansion of the external potentials such that they differ by more than a constant $C(t)$. It therefore follows that there exists some n such that the right hand side of the above equation (4.9) does not vanish. From this it can be concluded that the currents of the two systems defined by external potentials $V_{\mathrm{ext}}(\mathbf{r},t)$ and $V'_{\mathrm{ext}}(\mathbf{r},t)$ and evolving from the same initial states Ψ_0 must differ at a time $t_0+\delta$, where δ is some infinitesimal number.

Equation (4.9) thus guarantees that the two potentials $V_{\mathrm{ext}}(\mathbf{r},t)$ and $V'_{\mathrm{ext}}(\mathbf{r},t)$ must produce different density currents when evolving from the same initial state. However, the differences in the currents and the time-dependent densities of the two systems are related through a continuity equation of the form

$$\frac{\partial}{\partial t}\left[\rho(\mathbf{r},t)-\rho'(\mathbf{r},t)\right]=-\nabla\cdot\left[\mathbf{j}(\mathbf{r},t)-\mathbf{j}'(\mathbf{r},t)\right]. \tag{4.10}$$

Taking the $(n+1)$th time derivative of the above equation evaluated at time $t=t_0$ and substituting the result from equation (4.9) one obtains

$$\frac{\partial^{n+2}}{\partial t^{n+2}}\left[\rho(\mathbf{r},t)-\rho'(\mathbf{r},t)\right]|_{t=t_0}=-\nabla\cdot\left[\rho(\mathbf{r},t_0)\nabla\left\{\frac{\partial^{n}}{\partial t^{n}}\left[V_{\mathrm{ext}}(\mathbf{r},t)-V'_{\mathrm{ext}}(\mathbf{r},t)\right]|_{t=t_0}\right\}\right] \tag{4.11}$$

and thus, a relation between the difference of the time-dependent densities at a time $t=t_0+\delta$ and the Taylor expansion of the difference of the external potentials at time $t=t_0$ has been established. If the right hand side of equation (4.11) does not vanish for some n, it can be concluded that the time-dependent densities of the two systems evolving from the same initial state differ an infinitesimal instant after the initial time t_0.

To show that the right hand side indeed cannot vanish if the assumption specified in equation (4.5) holds, a short hand notation is introduced:

$$\frac{\partial^{n}}{\partial t^{n}}\left[V_{\mathrm{ext}}(\mathbf{r},t)-V'_{\mathrm{ext}}(\mathbf{r},t)\right]|_{t=t_0}=u_n(\mathbf{r}). \tag{4.12}$$

One now assumes the contrary of the statement that is to be proven, ie. that

$$\nabla\cdot\left[\rho(\mathbf{r},t_0)\nabla u_n(\mathbf{r})\right]=0. \tag{4.13}$$

Integrating over all space and applying Green's identity, it can be found that

$$\int \mathrm{d}^3r\, u_n(\mathbf{r})\nabla\cdot[\rho(\mathbf{r},t_0)\nabla u_n(\mathbf{r})] = \frac{1}{2}\oint \rho(\mathbf{r},t_0)\left[\nabla u_n^2(\mathbf{r})\right]\cdot\mathrm{d}\mathbf{s} - \int \mathrm{d}^3r\,\rho(\mathbf{r},t_0)\left[\nabla u_n(\mathbf{r})\right]^2 = 0. \tag{4.14}$$

The first term on the right hand side is a surface integral that vanishes in finite systems, as $\rho(\mathbf{r},t_0)$ decays exponentially with distance[17]. Thus for the expression to hold it is required that the volume integral on the right hand side vanishes, and that therefore $\rho(\mathbf{r},t_0)\left[\nabla u_n(\mathbf{r})\right]^2 = 0$. This however causes a contradiction with the original statement in equation (4.5), since the statement requires that $u_n(\mathbf{r})$ is not constant in $\mathbf{r}$ for some n [69]. Since both $\rho(\mathbf{r},t_0)$ and $\left[\nabla u_n(\mathbf{r})\right]^2$ are therefore greater or equal to zero everywhere in space, an integral of the product of them taken over all space cannot vanish. It can therefore be concluded that the statement made in equation (4.13) must be false and that time-dependent densities of systems with different external potentials evolving from the same initial state Ψ_0 are different for any $t > t_0$. From this, it follows directly that there exists a one-to-one mapping between the time-dependent density and the external potential (up to the scalar function $C(t)$) for any many-electron system evolving from a fixed initial state Ψ_0.

From the first part of the Runge–Gross theorem it follows that the time-dependent many-body wavefunction $\Psi(t)$ must be defined by the time-dependent density of the system, up to a time dependent phase factor of the form $\tilde{\Psi}(t) = e^{-\mathrm{i}\alpha(t)}\Psi(t)$, where $(\mathrm{d}/\mathrm{d}t)\alpha(t) = C(t)$. It thus follows that the matrix element

$$\langle\Psi(t)|\mathrm{i}\partial/\partial(t) - \hat{H}(t)|\Psi(t)\rangle \tag{4.15}$$

must be a *unique* functional of the time-dependent density since the function $C(t)$ is cancelled exactly by the time derivative of the phase factor in the many-electron wavefunction [69]. Defining the quantum mechanical action integral of the many-body system as

$$A[\Psi] = \int_{t_0}^{t_1} \mathrm{d}t\,\langle\Psi(t)|\mathrm{i}\partial/\partial t - \hat{H}|\Psi(t)\rangle \tag{4.16}$$

it follows directly that A is a unique functional of the time-dependent density. Partitioning the Hamiltonian $\hat{H}$ into the system-independent, universal part $\hat{T} + \hat{V}_{\mathrm{ee}}$ and the system-

[17] In principle, requiring the electron density to vanish at infinity means that the proof of the Runge–Gross theorem presented here is not valid for infinite systems. This limitation is avoided in TDCDFT, as the mapping between the current density and external potential can be proven without relying on any quantity vanishing at infinity (see equation (4.9)).

dependent part consisting of the time-dependent external potential, one can write

$$A[\rho] = B[\rho] - \int_{t_0}^{t_1} \mathrm{d}t \int \mathrm{d}^3 r \, \rho(\mathbf{r}, t) V_{\mathrm{ext}}(\mathbf{r}, t) \tag{4.17}$$

with

$$B[\rho] = \int_{t_0}^{t_1} \mathrm{d}t \, \langle \Psi[\rho](t) | \mathrm{i}\partial/\partial t - \hat{T} - \hat{V}_{\mathrm{ee}} | \Psi[\rho](t) \rangle. \tag{4.18}$$

Since the operators $\hat{T}$ and $\hat{V}_{\mathrm{ee}}$ are the same for any N-particle system, $B[\rho]$ must be a universal functional for any N-particle system, thus proving the second part of the Runge–Gross theorem.

Therefore, the following statement can be made due to the two parts of the Runge–Gross theorem: since $\rho(\mathbf{r}, t)$ determines the many-electron wavefunction $\Psi(t)$ for a system specified by $V_{\mathrm{ext}}(\mathbf{r}, t)$ evolving from Ψ_0 up to a complex phase, it follows that the expectation value of any observable associated with a time-dependent operator must be a *unique* functional of $\rho(\mathbf{r}, t)$ and Ψ_0.

4.1.2 The time-dependent Kohn–Sham equations

The Runge–Gross theorem establishes the mapping between the time-dependent external potential and the time-dependent density of the many-particle system, thus requiring the many-electron wavefunction to be a functional of the density. However, this statement in itself is not particularly helpful. Just like in ground state DFT, the main power of the TDDFT approach is that it allows for a mapping of the many-electron system to a non-interacting system moving in an effective potential that produces the same time-dependent density. This mapping is derived in the original Runge–Gross paper [69] in the following way:

First, one considers the universal functional $S[\rho]$ such that

$$S[\rho] = \int_{t_0}^{t_1} \mathrm{d}t \, \langle \Psi[\rho](t) | \mathrm{i}\partial/\partial t - \hat{T} | \Psi[\rho](t) \rangle \tag{4.19}$$

is the kinetic energy part of the universal functional of the many-electron system. Now consider different functionals $S[\rho]$ for the case of different types of electron-electron interactions V_{ee}. In particular, let $S_0[\rho]$ denote the functional for a system of non-interacting electrons with the same time-dependent density of the interacting system. It is then

possible to write the universal functional $A[\rho]$ as

$$\begin{aligned} A[\rho] &= S_0[\rho] - \int_{t_0}^{t_1} \mathrm{d}t \int \mathrm{d}^3 r\, \rho(\mathbf{r},t) V_{\mathrm{ext}}(\mathbf{r},t) \\ &\quad -\frac{1}{2}\int_{t_0}^{t_1} \mathrm{d}t \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, \frac{\rho(\mathbf{r},t)\rho(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|} - A_{\mathrm{xc}}[\rho] \end{aligned} \tag{4.20}$$

where $A_{\mathrm{xc}}[\rho]$, analogously to the ground state DFT functional $E_{\mathrm{xc}}[\rho]$, contains all many-body contributions beyond the mean-field treatment of the Coulomb interaction. In particular, it can be written as

$$A_{\mathrm{xc}}[\rho] = S_0[\rho] - S[\rho] + \int_{t_0}^{t_1} \mathrm{d}t\, \langle \Psi[\rho](t)|\hat{V}_{\mathrm{ee}}|\Psi[\rho](t)\rangle \tag{4.21}$$

$$-\frac{1}{2}\int_{t_0}^{t_1} \mathrm{d}t \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, \frac{\rho(\mathbf{r},t)\rho(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}. \tag{4.22}$$

As in ground-state DFT, the above is just a rewriting of the original action functional and the functional $A_{\mathrm{xc}}[\rho]$ is generally unknown. However, if it can be approximated in some way, the stationary point of the action functional in (4.20) can then be used to specify the fictitious non-interacting system. Taking the functional derivative of $A[\rho]$ one obtains

$$\frac{\delta A[\rho]}{\delta \rho(\mathbf{r},t)} = 0 = \frac{\delta S_0}{\delta \rho(\mathbf{r},t)} - V_{\mathrm{KS}}(\mathbf{r},t) \tag{4.23}$$

where

$$V_{\mathrm{KS}}(\mathbf{r},t) = V_{\mathrm{ext}}(\mathbf{r},t) + \int \mathrm{d}^3 r'\, \frac{\rho(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta A_{\mathrm{xc}}[\rho]}{\delta \rho(\mathbf{r},t)} \tag{4.24}$$

Note that Equation (4.23) is the Euler-Lagrange equation for a system of independent particles moving in the time-dependent effective potential $V_{\mathrm{KS}}(\mathbf{r},t)$. One can thus introduce $N/2$ single-particle orbitals $\{\psi_j^{\mathrm{KS}}\}$ following the effective single particle time-dependent Schrödinger equation

$$\left(\mathrm{i}\frac{\partial}{\partial t} + \frac{1}{2}\nabla^2\right)\psi_j^{\mathrm{KS}}(\mathbf{r},t) = V_{\mathrm{KS}}(\mathbf{r},t)\psi_j^{\mathrm{KS}}(\mathbf{r},t). \tag{4.25}$$

The time-dependent density of the interacting system can then be written in terms of the non-interacting Kohn–Sham states, such that

$$\rho(\mathbf{r},t) = 2\sum_j^{N/2} \psi_j^{\mathrm{KS}*}(\mathbf{r},t)\psi_j^{\mathrm{KS}}(\mathbf{r},t). \tag{4.26}$$

Note that just like in ground-state DFT, the Kohn–Sham potential is an explicit functional

of the density.

As mentioned before, $\delta A_{\mathrm{xc}}[\rho]/\delta\rho(\mathbf{r},t) = V_{\mathrm{xc}}(\mathbf{r},t)$ needs to be approximated in any practical calculation, just like in ground state DFT an approximation of the functional derivative $\delta E_{\mathrm{xc}}/\delta\rho(\mathbf{r})$ is needed. Note however, that finding suitable approximations to $V_{\mathrm{xc}}(\mathbf{r},t)$ at a given point $t_1 > t_0$ is more challenging than in the ground state DFT case. The reason for this can be found in the functional form of $A_{\mathrm{xc}}[\rho]$, which suggests that $V_{\mathrm{xc}}(\mathbf{r},t_1)$ is a functional of both the density $\rho(\mathbf{r},t)$ at all past times $t_0 < t \leq t_1$, and the initial many-body wavefunction Ψ_0 from which the system evolves.

The initial state and memory problem are discussed in more detail in a later section, but it is obvious that finding approximate functionals for TDDFT is more problematic than in the ground-state DFT case. Most approximations to the exchange-correlation functional in TDDFT thus rely on avoiding the memory dependence entirely and are referred to as adiabatic approximations. The simplest flavour, the ALDA or adiabatic LDA functional, simply states that

$$V_{\mathrm{xc}}^{\mathrm{ALDA}}(\mathbf{r},t) = \left.\frac{\delta E_{\mathrm{xc}}^{\mathrm{LDA}}}{\delta\rho(\mathbf{r})}\right|_{\rho=\rho(\mathbf{r},t)} \tag{4.27}$$

and thus the exchange-correlation potential at time t_1 is the same as the DFT exchange-correlation potential for a density $\rho = \rho(\mathbf{r},t_1)$. Since $V_{\mathrm{xc}}^{\mathrm{ALDA}}(\mathbf{r},t)$ only depends on the instantaneous time-dependent density at t_1, all memory effects, as well as any dependence on the initial state Ψ_0 are ignored in the ALDA functional.

4.1.3 Some comments on the Runge–Gross theorem

The original proof of the Runge–Gross theorem contains a number of technical problems. Some of them were explicitly mentioned in the original paper [69], while others were only discovered at a later point. This section contains a brief overview of some of these points and discusses attempts that have been made to improve the theoretical foundations of TDDFT.

One issue with the Runge–Gross approach to deriving the time-dependent Kohn–Sham equations can be seen in the functional derivative of the action functional $A[\rho]$. To see why this derivative is somewhat problematic from a theoretical point of view, it needs to be appreciated that functionals like $A[\rho]$ and $B[\rho]$ are not functionals of any arbitrary densities, but only densities that correspond to the time-dependent electron density of some system evolving from Ψ_0 in some external potential $V_{\mathrm{ext}}(\mathbf{r},t)$. That is, functionals $A[\rho]$ and $B[\rho]$ are only functionals of v-representable densities and thus the functional derivative in equation (4.23) can not be carried out over arbitrary changes in the density. The problem of v-representability is avoided in ground-state DFT by introducing the

Levy constrained search approach (see section 3.1.2). However, that approach relies on the variational principle and is thus only applicable in the ground state case. There has been some effort into extending the Runge–Gross theorem in order to allow for functionals that are valid for N-representable densities [70,72], but these approaches are more involved than the Levy constrained search approach in ground state DFT and will not be discussed here.

A second issue can be seen in the fact that all functionals in the Runge–Gross theorem have an implicit dependence on the many-electron wavefunction Ψ_0 from which the system initially evolves. Thus the Runge–Gross theorem does not formally replace all references to many-electron wavefunctions by references to the much simpler many-electron density. From a conceptual point of view, a removal of the functional dependence on the initial state Ψ_0 can be achieved by forcing the initial wavefunction to be the ground state of a system with external potential $V_{\text{ext}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}, t_0)$. In this case, the Hohenberg–Kohn theorem described in section 3.1.1 guarantees that $\Psi_0 = \Psi_0[\rho_0]$, and thus can be written as a functional of the ground state density of the system specified by the external potential at t_0. Since the main interest of this work is the calculation of excitations caused by perturbations to a system in its ground state, the dependence of $A[\rho]$ on the initial many-body wavefunction can be removed for this class of problems.

Another point that can be raised concerning the Runge–Gross theorem is its dependence on the existence of a Taylor expansion of $V_{\text{ext}}(\mathbf{r}, t)$ at time $t = t_0$. In other words, switching on the external time-dependent potential instantaneously at t_0 or considering it to be a short a pulse in the form of a delta function at t_0 is not strictly valid in the framework of the Runge–Gross theorem.

A number of issues with the original Runge–Gross theorem have to do with the functional form of the action used to derive the time-dependent Kohn–Sham equations. In the original paper by Runge and Gross, the time-dependent Schrödinger equation can only be derived from the action functional when treating the boundary conditions in such a way that $\delta\Psi(t_0) = \delta\Psi(t_1) = 0$, in other words the initial and final many-electron states of the propagation have to be treated as fixed. However, since the Schrödinger equation is a first order differential equation in time, any variation $\delta\Psi(t)$ at $t > t_0$ is completely and uniquely determined by fixed initial condition $\delta\Psi(t_0) = 0$, meaning that the required boundary condition at $t = t_1$ is not free to be specified [73]. Furthermore, it is found that the original formulation of the action functional does violate causality when considering response functions like $\delta V_{\text{xc}}(\mathbf{r}, t)/\delta\rho(\mathbf{r}', t')$, which are required to vanish for $t < t'$ but are instead found to be symmetric in time in the original formulation. These problems are addressed by van Leeuwen [73] in an influential paper that puts TDDFT on more rigorous foundations using techniques derived from the treatment of non-equilibrium many-body

Green's functions.

In summary, the theoretical framework of TDDFT is considerably more subtle than the foundations of DFT and is an area of ongoing research.

4.2 From theory to practical method

While the theoretical foundations of TDDFT are presented in the the previous section, the focus is now shifted to introducing the basics of TDDFT as a practical method for calculating the low energy optical response of a given system. Two main flavours of TDDFT, the real-time propagation of the Kohn–Sham equations and the linear-response frequency domain treatment, will be discussed in the light of the objective of calculating low energy excitations in very large systems. The role of non-locality and memory effects in the response of the TDDFT exchange-correlation potential to perturbations in the density is then discussed in some greater detail.

4.2.1 The real-time TDDFT approach

The first approach to calculating absorption spectra is to extract the required information directly from the explicit time evolution of the Kohn–Sham system. The task is to evaluate the photoabsorption cross-section $\sigma(\omega)$ given by [74]

$$\sigma(\omega) = \frac{4\pi\omega}{3} \sum_{\gamma} \mathrm{Im}\left[\delta_{\gamma,\nu}\alpha_{\gamma\nu}(\omega)\right], \tag{4.28}$$

where $\alpha_{\gamma\nu}(\omega)$ is the dynamic polarisability tensor and the labels γ, $\nu = x, y, z$ denotes the direction of polarisation. In the dipole approximation and in the limit of weak perturbation, $\alpha_{\gamma\nu}(\omega)$ is related to the linear density response function $\chi(\mathbf{r}, \mathbf{r}', \omega)$ via

$$\alpha_{\gamma\nu}(\omega) = -\int \mathrm{d}^3r\, \mathrm{d}^3r'\, r_\gamma \chi(\mathbf{r}, \mathbf{r}', \omega) r'_\nu \tag{4.29}$$

The linear density response function is then defined by the change of the density due to the perturbation of an infinitesimal external potential $\delta V_{\mathrm{pert}}(\mathbf{r}, t)$. Thus, the induced perturbation of the density of the system can be written as

$$\rho^{\{1\}}(\mathbf{r}, \omega) = \int \mathrm{d}^3r'\, \chi(\mathbf{r}, \mathbf{r}', \omega) \delta V_{\mathrm{pert}}(\mathbf{r}', \omega) \tag{4.30}$$

Consider a Kohn–Sham system with external potential $V_{\mathrm{ext}}(\mathbf{r})$ in its ground state at $t = t_0$, with an associated ground state electron density of $\rho_0(\mathbf{r})$. At $t = t_0^+$, the system is perturbed by the external potential $\delta V_{\mathrm{pert}}(\mathbf{r}, t) = -E_{\mathrm{ext}} r_\mu \delta(t - t_0)$, where E_{ext} is chosen

small enough to ensure that the system response is in the linear regime [75]. It then follows directly from equation (4.29) that the dynamical polarisability can be expressed through the first order change in density $\rho^{\{1\}}(\mathbf{r}, \omega)$ via

$$\alpha_{\gamma\nu}(\omega) = \frac{1}{E_{\text{ext}}} \int \mathrm{d}^3 r \, r_\gamma \rho^{\{1\}}(\mathbf{r}, \omega). \tag{4.31}$$

Note however, that the time-dependent density $\rho(\mathbf{r}, t)$ of the system is directly accessible from solving the time dependent Kohn–Sham equations. Therefore, the change in density $\rho^{\{1\}}(\mathbf{r}, \omega)$ can be trivially obtained from a Fourier transform of $\rho^{\{1\}}(\mathbf{r}, t) = \rho(\mathbf{r}, t) - \rho_0(\mathbf{r})$ and thus the absorption spectrum of the system can be directly computed from the time-dependent density.

The remaining issue is then how to efficiently solve the time-dependent Kohn–Sham equations to obtain the Kohn–Sham orbitals and the density evolving from the initial ground state of the system. This generally requires approximation of the time evolution operator $\hat{U}(t + \delta t, t)$, that propagates the Kohn–Sham orbitals at time t to some later time $t + \delta t$, where δt is a chosen time step. A number of different approximations have been proposed, and a survey of some common methods can be found in [76]. Here it should just be mentioned that the time step size has to be picked in such a way that the propagation remains stable, as approximate propagators are not strictly unitary and do not exactly conserve the norm of the Kohn–Sham states.

Since any motion of the ions and thus any coupling of excitations to phonons is ignored, the spectrum that a real time TDDFT calculation yields does in fact correspond to a number of delta functions at frequencies equal to the excitation energies of the system. However, this limit is only achieved in an infinite resolution, which can only be obtained with an infinite time propagation of the orbitals and an exact propagator. Thus, a small imaginary frequency is introduced into the Fourier transform of the time-dependent density response of the system, which results in a smearing out of excitation peaks and a smoothing of the spectrum. This broadening term, while being purely introduced for computational convenience, can be seen as approximating the effect of electron-phonon coupling on the system and makes spectra more comparable to experimental results.

4.2.2 The linear response formalism

In the previous section it is shown how the time-dependent density response of the system can yield the entire absorption spectrum. However, in many situations, one is mainly interested in the excitation energies of a few low-energy excitations, rather than an entire spectrum. In these situations, propagating the entire Kohn–Sham system in time can become cumbersome and ideally a method should be chosen that avoids the explicit time

propagation. This method is the frequency-domain linear-response TDDFT formalism.

It is worth pointing out that the term "linear-response formalism" is to a certain extent misleading, since one way to obtain the absorption spectrum from the time-dependent density outlined in the previous section also assumes a linear response of the system to an external perturbation. However, since the time propagation of the real-time TDDFT scheme can be carried out without assuming a linear response and can thus be used to treat the non-linear regime as well, the method that is generally referred to as linear-response TDDFT is taken to be a purely frequency-domain linear response scheme that does not require any explicit propagation of Kohn–Sham states.

In order to avoid the explicit propagation of the Kohn–Sham orbitals, the linear density response $\chi(\mathbf{r}, \mathbf{r}', \omega)$ of the system is studied in some more detail. Here, one again considers a Kohn–Sham system with external potential $V_{\text{ext}}(\mathbf{r})$ in its ground state at $t = t_0$, with associated ground state density $\rho_0(\mathbf{r})$. The physical quantity of interest is the response of the system to an infinitesimal external time-dependent perturbation $\delta V_{\text{pert}}(\mathbf{r}, t)$, such that the total external potential of the system is given by $V_{\text{ext}}(\mathbf{r}, t) = V_{\text{ext}}(\mathbf{r}) + \delta V_{\text{pert}}(\mathbf{r}, t)$.

In real space and time, the response function can be expressed as [77, 78]

$$\chi(\mathbf{r}, t, \mathbf{r}', t') = \left.\frac{\delta\rho(\mathbf{r}, t)}{\delta V_{\text{ext}}(\mathbf{r}', t')}\right|_{V_{\text{ext}}[\rho_0]} \tag{4.32}$$

where the functional derivative is evaluated at $V_{\text{ext}}[\rho_0](\mathbf{r}, t) = V_{\text{ext}}(\mathbf{r})$. Thus $\chi(\mathbf{r}, t, \mathbf{r}', t')$ measures the first order change of the time-dependent density of the interacting system due to a small perturbation to the external potential of the system. However, one can similarly consider the density response function $\chi_0(\mathbf{r}, t, \mathbf{r}', t')$ of the fictitious non-interacting Kohn–Sham system moving in an external time-dependent potential $V_{\text{KS}}(\mathbf{r}, t)$ (see Eqn. (4.24)), which is given by

$$\chi_0(\mathbf{r}, t, \mathbf{r}', t') = \left.\frac{\delta\rho(\mathbf{r}, t)}{\delta V_{\text{KS}}(\mathbf{r}', t')}\right|_{V_{\text{KS}}[\rho_0]}. \tag{4.33}$$

Here, the functional derivative is evaluated at $V_{\text{KS}}[\rho_0]$, corresponding to the static ground state Kohn–Sham potential. Since the time-dependent density generated by the fictitious Kohn–Sham system and the full interacting system must be identical due to the Runge–Gross theorem, it is then possible to link the the two response functions via the chain rule of differentiation [78]:

$$\chi(\mathbf{r}, t, \mathbf{r}', t') = \int \mathrm{d}^3 r'' \int \mathrm{d}t'' \left.\frac{\delta\rho(\mathbf{r}, t)}{\delta V_{\text{KS}}(\mathbf{r}'', t'')} \frac{\delta V_{\text{KS}}(\mathbf{r}'', t'')}{\delta V_{\text{ext}}(\mathbf{r}', t')}\right|_{V_{\text{ext}}[\rho_0]}. \tag{4.34}$$

The functional derivative of the Kohn–Sham potential with respect to the external po-

tential can be straightforwardly evaluated, yielding

$$\left.\frac{\delta V_{\mathrm{KS}}(\mathbf{r}'',t'')}{\delta V_{\mathrm{ext}}(\mathbf{r}',t')}\right|_{V_{\mathrm{ext}}[\rho_0]} = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') + \int \mathrm{d}^3r'' \int \mathrm{d}\,t'' \left[\frac{\delta(t-t'')}{|\mathbf{r}-\mathbf{r}''|} + \left.\frac{\delta A_{\mathrm{xc}}}{\delta\rho(\mathbf{r},t)\delta\rho(\mathbf{r}'',t'')}\right|_{V_{\mathrm{KS}}[\rho_0]}\right] \chi(\mathbf{r}'',t'',\mathbf{r}',t') \quad (4.35)$$

It is now convenient to introduce a short hand notation for the second derivative of the exchange-correlation density functional, evaluated at the ground-state Kohn–Sham potential, such that

$$f_{\mathrm{xc}}(\mathbf{r},t,\mathbf{r}',t') = \left.\frac{\delta^2 A_{\mathrm{xc}}}{\delta\rho(\mathbf{r},t)\delta\rho(\mathbf{r}',t')}\right|_{V_{\mathrm{KS}}[\rho_0]}. \quad (4.36)$$

The function $f_{\mathrm{xc}}(\mathbf{r},t,\mathbf{r}',t')$ is of major importance in the linear-response TDDFT scheme and is commonly referred to as the exchange-correlation kernel. It should be noted that, in line with comments made in section 4.1.3, the definition of $f_{\mathrm{xc}}(\mathbf{r},t,\mathbf{r}',t')$ as a second functional derivative of A_{xc} is not formally correct, since it assumes a symmetry under the exchange of t and t' and thus a violation of causality. However, since in practice the exchange-correlation kernel has to be approximated in any case it is simply pointed out that the exact conditions that $f_{\mathrm{xc}}(\mathbf{r},t,\mathbf{r}',t')$ needs to fulfil cannot be derived from equation (4.36).

Using Equation (4.34) and (4.35), it is then possible to define a Dyson equation in frequency domain for the fully interacting response function $\chi(\mathbf{r},\mathbf{r}',\omega)$:

$$\chi(\mathbf{r},\mathbf{r}',\omega) = \chi_0(\mathbf{r},\mathbf{r}',\omega) + \int \mathrm{d}^3r_1\,\mathrm{d}^3r_2\,\chi_0(\mathbf{r},\mathbf{r}_1,\omega)\left[\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|} + f_{\mathrm{xc}}(\mathbf{r}_1,\mathbf{r}_2,\omega)\right]\chi(\mathbf{r}_2,\mathbf{r}',\omega) \quad (4.37)$$

The above equation can be solved self-consistently for the full density response $\chi(\mathbf{r},\mathbf{r}',\omega)$, starting from the density response of the fictitious Kohn–Sham system. Furthermore, equation (4.37) can then be used to find an expression for the perturbation of the density $\rho^{\{1\}}(\mathbf{r},\omega)$ in terms of the non-interacting Kohn–Sham system only, where one can write [78]

$$\rho^{\{1\}}(\mathbf{r},\omega) = \int \mathrm{d}^3r'\,\chi_0(\mathbf{r},\mathbf{r}',\omega) \times \left[\delta V_{\mathrm{pert}}(\mathbf{r}',\omega) + \int \mathrm{d}^3r_1\,\left(\frac{1}{|\mathbf{r}'-\mathbf{r}_1|} + f_{\mathrm{xc}}(\mathbf{r}',\mathbf{r}_1,\omega)\right)\rho^{\{1\}}(\mathbf{r}_1,\omega)\right]. \quad (4.38)$$

Equation (4.38) defines a self-consistency relation for the frequency-dependent density response of the system purely in terms of quantities related to the fictitious non-interacting Kohn–Sham system. From Equation (4.31) it is known that $\rho^{\{1\}}(\mathbf{r},\omega)$ contains all infor-

mation required to construct the absorption spectrum of a system. Note that if f_{xc} is fully known, equation (4.38) is guaranteed to deliver the *exact* linear density response of the interacting many-electron system thanks to the Runge–Gross theorem.

Since $\chi_0(\mathbf{r}, \mathbf{r}', \omega)$ is fully characterised by the ground state of the unperturbed Kohn–Sham system at $t = t_0$, and the exchange-correlation kernel $f_{\mathrm{xc}}(\mathbf{r}, \mathbf{r}', \omega)$ is evaluated at the ground state density ρ_0, the absorption spectrum can be calculated from the ground state of the unperturbed Kohn–Sham system only by solving equation (4.38). Thus, no explicit time propagation of Kohn–Sham orbitals is necessary in order to calculate the linear density response of the system.

To calculate the density response of the system, one has to introduce an explicit representation of $\chi_0(\mathbf{r}, \mathbf{r}', \omega)$. Let $\{\psi_i^{\mathrm{KS}}(\mathbf{r})\}$ denote the set of all ground state Kohn–Sham orbitals with eigenenergies $\{\epsilon_i^{\mathrm{KS}}\}$, corresponding to the system with external potential $V_{\mathrm{ext}}(\mathbf{r})$ and ground state density ρ_0. Furthermore, let n_k denote a general occupation number such that $n_k = 0$ for $k > N_{\mathrm{occ}}/2$ and $n_k = 1$ otherwise. Then, the Kohn–Sham response function can simply be written as

$$\chi_0(\mathbf{r}, \mathbf{r}', \omega) = 2 \sum_{j,k=1}^{\infty} (n_k - n_j) \frac{\psi_k^{\mathrm{KS}*}(\mathbf{r}) \psi_j^{\mathrm{KS}}(\mathbf{r}) \psi_j^{\mathrm{KS}*}(\mathbf{r}') \psi_k^{\mathrm{KS}}(\mathbf{r}')}{\omega - (\epsilon_k^{\mathrm{KS}} - \epsilon_j^{\mathrm{KS}}) + \mathrm{i}\delta} \tag{4.39}$$

where the infinitesimal imaginary part $\mathrm{i}\delta$ is added in order to ensure the convergence of the Fourier transform of $\chi_0(\mathbf{r}, \mathbf{r}', \omega)$ into the time domain. From equation (4.39), it can be seen that χ_0 only contains terms that correspond to particle-hole (for $n_j = 1$ and $n_k = 0$) or hole-particle (for $n_k = 1$ and $n_j = 0$) Kohn–Sham transitions and that it diverges for ω equal to a Kohn–Sham particle-hole or hole particle transition energy $\epsilon_{ij}^{\mathrm{KS}} = \epsilon_i^{\mathrm{KS}} - \epsilon_j^{\mathrm{KS}}$. Thus, if one uses equation (4.38) to directly calculate the density response and the spectrum, it becomes necessary for numerical reasons to add a small imaginary damping term to the frequency in order to avoid the poles along the real axis. Just like in the time-domain TDDFT approach this has the effect of removing any singularities and smoothing out the spectrum, and can be interpreted physically as an approximate treatment of electron-phonon coupling (ie. temperature) effects on the spectrum.

4.2.3 The exchange-correlation kernel

Before moving on in the derivation of an equation that yields the individual excitation energies directly, without the need to construct the density response of the system for the entire frequency range, it is convenient to discuss the nature of the exchange-correlation kernel in some detail. As in the case of time-domain TDDFT, where the first functional derivative of the exchange-correlation density functional A_{xc} has to be approximated, f_{xc} is generally unknown and suitable approximations have to be found in order to perform

practical calculations.

Note however, that while $V_{\mathrm{xc}}(\mathbf{r},t) = \delta A_{\mathrm{xc}}[\rho]/\delta\rho(\mathbf{r},t)$ is an explicit functional of the time-dependent density and is dependent on the value of $\rho(\mathbf{r},t_1)$ at all previous values $t_1 < t$, the functional derivative in f_{xc} is evaluated at the ground-state density of the system only, and is thus uniquely specified through ρ_0.

The true TDDFT kernel f_{xc} is required to be fully non-local and frequency dependent, accounting for all dynamical exchange and correlation effects in the response of the system to an external perturbation that go beyond the simple mean-field treatment of the electrons. It is thus inherently difficult to approximate. As a first step, one often makes the approximation of ignoring the frequency dependence of $f_{\mathrm{xc}}(\mathbf{r},\mathbf{r}',\omega)$, thus removing any dynamical exchange-correlation effects from the calculation. Approximating the exchange correlation kernel as static is, in accordance with the real time TDDFT approach, known as the adiabatic approximation. The simplest available approximation to the TDDFT kernel is then the adiabatic LDA approximation, where

$$f_{\mathrm{xc}}^{\mathrm{ALDA}}(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')\left.\frac{\partial V_{\mathrm{xc}}^{\mathrm{LDA}}[\rho]}{\partial\rho(\mathbf{r})}\right|_{\rho=\rho_0} \tag{4.40}$$

and thus $f_{\mathrm{xc}}^{\mathrm{ALDA}}$ is simply obtained from the derivative of the ground state Kohn–Sham exchange-correlation potential evaluated at the ground state density. Similarly, adiabatic GGA functionals can be obtained by taking the ground state Kohn–Sham exchange correlation potential in the GGA approximation. Note that for any semi-local DFT exchange correlation functional, the required non-locality of f_{xc} derived from it is automatically lost. Thus in the ALDA and similar approximations, f_{xc} is taken to be fully local and static. In section 4.2.6, the discussion will focus in some more detail on the consequences these common approximations have on the excitation energies and spectra derived from TDDFT.

4.2.4 The effective 2-particle eigenvalue equation

Equation (4.38) establishes the self-consistent linear density response of the system as a reaction to an external perturbation $\delta V_{\mathrm{pert}}(\mathbf{r},\omega)$ in terms of the non-interacting linear response function χ_0. The linear density response $\rho^{\{1\}}(\mathbf{r},\omega)$ is linked to the linear response function $\chi(\omega)$ of the fully interacting system via equation (4.30). Note however, that from the considerations in section 4.2.1 it follows that $\chi(\omega)$ has poles at the true excitation energies of the system. Thus finding the position of the poles directly yields the excitation energies of the system.

As a first step, one substitutes the expression for $\chi_0(\omega)$ in terms of unperturbed ground state Kohn–Sham orbitals into equation (4.38). It is then possible to write the equation

in matrix form, by expressing all spatial dependencies in the basis of ground state Kohn–Sham orbitals. After a few rearrangements, one obtains[18] [79]:

$$\sum_{kl}^{n_k-n_l\neq 0}\left[\delta_{i,k}\delta_{j,l}\frac{\omega-(\epsilon_k^{\mathrm{KS}}-\epsilon_l^{\mathrm{KS}})}{n_l-n_k}-K_{ij,kl}(\omega)\right]\rho_{kl}^{\{1\}}(\omega)=\delta V_{ij}^{\mathrm{pert}}(\omega) \tag{4.41}$$

where

$$\rho_{ij}^{\{1\}}(\omega)=\int \mathrm{d}^3r\,\psi_i^{\mathrm{KS}*}(\mathbf{r})\rho^{\{1\}}(\mathbf{r},\omega)\psi_j^{\mathrm{KS}}(\mathbf{r}) \tag{4.42}$$

$$\delta V_{ij}^{\mathrm{pert}}(\omega)=\int \mathrm{d}^3r\,\psi_i^{\mathrm{KS}*}(\mathbf{r})\delta V_{\mathrm{pert}}(\mathbf{r},\omega)\psi_j^{\mathrm{KS}}(\mathbf{r}) \tag{4.43}$$

$$K_{ij,kl}=\int \mathrm{d}^3r\,\mathrm{d}^3r'\,\psi_i^{\mathrm{KS}*}(\mathbf{r})\psi_j^{\mathrm{KS}}(\mathbf{r})\left[\frac{1}{|\mathbf{r}-\mathbf{r}'|}+f_{\mathrm{xc}}(\mathbf{r},\mathbf{r}',\omega)\right]\psi_k^{\mathrm{KS}}(\mathbf{r}')\psi_l^{\mathrm{KS}*}(\mathbf{r}') \tag{4.44}$$

One can then consider the structure of the above matrix equation with indices i, j, k and l, in order to separate particle-hole and hole-particle like transitions. Let c denote the label of an unoccupied Kohn–Sham state and v the label of an occupied Kohn–Sham state, such that $\rho_{cv}^{\{1\}}$ then corresponds to a particle-hole matrix element and $\rho_{vc}^{\{1\}}$ to a hole-particle one. Considering hole-particle and particle-hole transitions individually and using the fact that the coupling matrix $K_{ij,kl}$ is invariant under the swapping of the i with the j, as well as the k with the l indices, equation (4.41) can be rewritten in terms of a block matrix equation [79, 80]:

$$\left\{\begin{pmatrix}\mathbf{A}(\omega) & \mathbf{B}(\omega)\\ -\mathbf{B}(\omega) & -\mathbf{A}(\omega)\end{pmatrix}-\omega\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{pmatrix}\right\}\begin{pmatrix}\vec{\rho}^{\{1\}}(\omega)\\ \vec{\rho}^{\{1\}}(\omega)\end{pmatrix}=\begin{pmatrix}-\delta\mathbf{V}^{\mathrm{pert}}(\omega)\\ \delta\mathbf{V}_{\mathrm{pert}}(\omega)\end{pmatrix} \tag{4.45}$$

where[19]

$$A_{cv,c'v'}(\omega) = \delta_{cc'}\delta_{vv'}(\epsilon_{c'}^{\mathrm{KS}}-\epsilon_{v'}^{\mathrm{KS}})+2K_{cv,c'v'}(\omega) \tag{4.46}$$

$$B_{cv,c'v'}(\omega) = 2K_{cv,v'c'}(\omega). \tag{4.47}$$

Thus, the matrix equation (4.45) consists of diagonal blocks corresponding to particle-hole and hole-particle transitions, while the off-diagonal blocks $\mathbf{B}(\omega)$ couple particle-hole and hole-particle transtions. Physically, particle-hole and hole-particle transitions can be interpreted as excitations and de-excitations and the coupling between them is a consequence of time-reversal symmetry in the system. Now note that for frequencies ω

[18] The quantity in brackets linking $\rho_{ij}^{\{1\}}(\omega)$ and $\delta V_{ij}^{\mathrm{pert}}(\omega)$ actually corresponds to the inverse of $\chi(\omega)$ in matrix form.

[19] Here, the factor of 2 originates from an implied summation over spin indices. The spin structure of the equation is discussed in some more detail in section 4.2.5.

equal to an excitation energy of the system, it is required that the dynamic polarisability of the system has a pole, which can only occur if the block matrix on the left hand side of equation (4.45) is non-invertible [77]. From this it follows that excitation energies can be directly obtained by solving the effective eigenvalue problem [79, 80]

$$\mathbf{H}^{[2p]}(\omega)\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \omega\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \tag{4.48}$$

where

$$\mathbf{H}^{[2p]}(\omega) = \begin{pmatrix}\mathbf{A}(\omega) & \mathbf{B}(\omega)\\ -\mathbf{B}(\omega) & -\mathbf{A}(\omega)\end{pmatrix} \tag{4.49}$$

Note that solutions to equation (4.48) are generally not trivially obtained, since $\mathbf{H}^{[2p]}$ explicitly depends on ω. However, if the ω-dependence of the exchange-correlation kernel is ignored, solving (4.48) simply corresponds to solving a non-Hermitian eigenvalue problem. Solutions then correspond to all particle-hole and hole-particle excitations in the system. It was however noted by Casida [79], that the above eigenvalue problem can be transformed into a simpler one of the form

$$\Omega(\omega)\mathbf{F} = \omega^2\mathbf{F} \tag{4.50}$$

in terms of the pseudoeigenvectors $\mathbf{F}$. $\Omega(\omega)$ is now only defined in particle-hole space and is given by

$$\Omega(\omega) = \delta_{cc'}\delta_{vv'}(\epsilon^{\mathrm{KS}}_{c'} - \epsilon^{\mathrm{KS}}_{v'})^2 + 2\sqrt{(\epsilon^{\mathrm{KS}}_{c} - \epsilon^{\mathrm{KS}}_{v})}K_{cv,c'v'}(\omega)\sqrt{(\epsilon^{\mathrm{KS}}_{c'} - \epsilon^{\mathrm{KS}}_{v'})} \tag{4.51}$$

The reformulation of the two-particle Hamiltonian eigenvalue equation in terms of $\Omega(\omega)$ is known as Casida's equation and allows for the direct solution for the square of the excitation energies. However, the reformulation requires the evaluation of a matrix $\mathbf{S}$ that contains the square root of Kohn–Sham eigenvalue differences in its diagonal entries. While in Kohn–Sham particle-hole representation this matrix is trivially calculated, it cannot be easily obtained if another representation of the TDDFT 2-particle Hamiltonain eigenvalue matrix is chosen and is thus strictly limited to the representation considered here.

Assuming that there is no frequency dependence in $\Omega(\omega)$, the mean polarisability $\alpha(\omega)$ can then be written in terms of excitation energies ω_i and the oscillator strengths f_i of excitations expressible through the normalised pseudoeigenvectors $\mathbf{F}$ such that

$$\alpha(\omega) = \sum_i \frac{f_i}{\omega_i^2 - \omega^2} \tag{4.52}$$

and

$$f_i = \frac{2}{3} \sum_{\gamma=x,y,z} \left| \sum_{c',v} \langle \psi_v^{\mathrm{KS}} | r_\gamma | \psi_{c'}^{\mathrm{KS}} \rangle \sqrt{(\epsilon_{c'}^{\mathrm{KS}} - \epsilon_v^{\mathrm{KS}})} F_{c'v} \right|^2 \tag{4.53}$$

There is an important sum rule (Thomas–Reiche–Kuhne) [81] connected with the oscillator strengths $\{f_i\}$, in that their sum must equal the number of electrons in the system. In the derivation so far, it has been assumed that the matrix equations derived from (4.41) are expanded in an infinite set of Kohn–Sham eigenstates of the unperturbed system. In practice, these Kohn–Sham states are represented in an underlying basis and thus only a finite number of states is used. The sum rule can then be used as a measure the quality of the basis set representation of the TDDFT eigenvalue equation[20].

4.2.5 The Tamm–Dancoff approximation

The structure of $\mathbf{H}^{[2p]}$ explicitly considers particle-hole and hole-particle excitations, as well as the coupling between them. Since this causes the full effective 2-particle Hamiltonian $\mathbf{H}^{[2p]}$ to be non-Hermitian (see equation (4.48)), and since non-Hermitian eigenvalue problems are cumbersome to solve using computational methods, it is desirable to find an approximation that casts equation (4.48) into fully Hermitian form. This can be easily achieved by the Tamm–Dancoff approximation (TDA) [82], in which the block matrices $\mathbf{B}(\omega)$ are ignored. Thus, any coupling between the particle-hole and hole-particle excitations is neglected and it is sufficient to solve for the particle-hole space only, with

$$\mathbf{H}_{\mathrm{TDA}}^{[2p]}(\omega)\mathbf{x} \equiv \mathbf{A}(\omega)\mathbf{x} = \omega\mathbf{x}. \tag{4.54}$$

Since the block matrix $\mathbf{A}(\omega)$ is trivially shown to be Hermitian, the above approximation can be solved with standard eigenvalue solvers for Hermitian eigenproblems. The Tamm–Dancoff approximation has been known to deliver excitation energies that are of comparable quality to the solution of the full TDDFT eigenvalue equation [83]. However, oscillator strengths derived from it are known to violate the Thomas–Reiche–Kuhne sum rule.

Up to this point any spin dependence of the effective TDDFT 2-particle Hamiltonian has been ingnored. Denoting the two different spin channels as $\uparrow$ and $\downarrow$ one can then divide the effective Hilbert space of the 2-particle Hamiltonian into four parts corresponding to all possible combinations of the spin channels of the particle and the hole: $c\uparrow v\uparrow$, $c\uparrow v\downarrow$, $c\downarrow v\uparrow$ and $c\downarrow v\downarrow$. The full Hamiltonian then decouples into two solution classes, a class of triplet solutions given by subspaces $c\uparrow v\downarrow$, $c\downarrow v\uparrow$ and $1/\sqrt{2}(c\uparrow v\uparrow + c\downarrow v\downarrow)$ and a

[20]Note that such a measure of basis set quality requires the calculation of all possible excitations and thus a full diagonalisation of the entire 2-particle Hamiltonian and is thus not practical for larger systems.

singlet class of solutions with effective subspace $1/\sqrt{2}(c\uparrow v\uparrow - c\downarrow v\downarrow)$ [84].

Provided the discussion is limited to closed-shell systems with spin-degenerate ground states only, the unperturbed ground state Kohn–Sham states $\{\psi_i^{\mathrm{KS}}\}$ and their associated eigenstates $\{\epsilon_i^{\mathrm{KS}}\}$ are spin-degenerate and thus the spin-dependence only enters in the coupling term $\mathbf{K}(\omega)$. Taking $f_{\mathrm{xc}}(\mathbf{r},\mathbf{r}',\omega)$ to be in the adiabatic approximation, such that it can be written as

$$f_{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = \left.\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\rho=\rho_0} \tag{4.55}$$

where E_{xc} is the exchange-correlation energy, it becomes obvious that the spin structure of $\mathbf{K}$ originates from the second order functional derivative with respect to the density. One can then differentiate between two different spin configurations for f_{xc}, with

$$f_{\uparrow\uparrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = f_{\downarrow\downarrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = \left.\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho_\uparrow(\mathbf{r})\delta\rho_\uparrow(\mathbf{r}')}\right|_{\rho=\rho_0} \tag{4.56}$$

$$f_{\uparrow\downarrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = f_{\downarrow\uparrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = \left.\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho_\uparrow(\mathbf{r})\delta\rho_\downarrow(\mathbf{r}')}\right|_{\rho=\rho_0}. \tag{4.57}$$

The singlet and triplet configurations can then be solved for by treating two different eigenvalue equations with different coupling matrices where the coupling matrices $\mathbf{K}^{\{\mathrm{triplet}\}}$ and $\mathbf{K}^{\{\mathrm{singlet}\}}$ for the triplet case and the singlet case can then be written as

$$\begin{aligned} K_{cv,c'v'}^{\{\mathrm{singlet}\}} &= \int \mathrm{d}^3r\,\mathrm{d}^3r'\,\psi_c^{\mathrm{KS}*}(\mathbf{r})\psi_v^{\mathrm{KS}}(\mathbf{r}) \\ &\quad\times\left[\frac{1}{|\mathbf{r}-\mathbf{r}'|} + 0.5\left(f_{\uparrow\uparrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}') + f_{\uparrow\downarrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}')\right)\right]\psi_{c'}^{\mathrm{KS}}(\mathbf{r}')\psi_{v'}^{\mathrm{KS}*}(\mathbf{r}') \end{aligned} \tag{4.58}$$

$$\begin{aligned} K_{cv,c'v'}^{\{\mathrm{triplet}\}} &= \frac{1}{2}\int \mathrm{d}^3r\,\mathrm{d}^3r'\,\psi_c^{\mathrm{KS}*}(\mathbf{r})\psi_v^{\mathrm{KS}}(\mathbf{r})\left(f_{\uparrow\uparrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}') - f_{\uparrow\downarrow}^{\mathrm{xc}}(\mathbf{r},\mathbf{r}')\right) \\ &\quad\times\psi_{c'}^{\mathrm{KS}}(\mathbf{r}')\psi_{v'}^{\mathrm{KS}*}(\mathbf{r}') \end{aligned} \tag{4.59}$$

Solving for the eigenstates of the two different 2-particle Hamiltonians $\mathbf{H}_{\{\mathrm{singlet}\}}^{[2p]}$ and $\mathbf{H}_{\{\mathrm{triplet}\}}^{[2p]}$ with coupling matrices $\mathbf{K}^{\{\mathrm{singlet}\}}$ and $\mathbf{K}^{\{\mathrm{triplet}\}}$ then yields the the triply degenerate triplet states and the set of singlet states of the system. It should be noted that the above analysis only considers spin-degenerate ground states and systems where any spin-orbit coupling of the electrons is negligible. If these two assumptions do not hold, a full treatment of the entire spin structure of the Hamiltonian $\mathbf{H}^{[2p]}$ becomes necessary and singlet and triplet subspaces are not easily decoupled [84].

4.2.6 Sources of errors in TDDFT calculations

In this section a practical formalism of calculating excitation spectra using TDDFT has been introduced, either by generating the entire spectrum from the explicit time-

dependent response density obtained from real-time TDDFT (RT-TDDFT), or by directly obtaining excitation energies as the solutions to an eigenvalue equation of an effective 2-particle Hamiltonian in linear-response TDDFT (LR-TDDFT). It is to be pointed out that both the RT-TDDFT approach and the LR-TDDFT method presented here are in principle exact (in the linear-response regime), that is if the exact ground state density and ground state Kohn–Sham potential, as well as the exact exchange-correlation potential as a functional of the time-dependent density are known, both methods necessarily produce the exact excitation spectrum of the many-electron system.

However, since both in the ground state DFT and the TDDFT scheme, the exchange-correlation functional has to be approximated, it is important to analyse the influence of these approximations on the quality of the excitation energies obtained from TDDFT. While the approximations have the same effect no matter whether the RT-TDDFT or the LR-TDDFT approach is used, the focus will be on the linear-response approach here, as the effective 2-particle Hamiltonian allows for a more intuitive analysis of some of the failings of commonly used approximate functionals.

The first obvious approximation introduced is that the ground state density ρ_0 of the Kohn–Sham system that is perturbed to derive the 2-particle Hamiltonian eigenvalue equation does not correspond to the true ground state density of the many-electron system, since the DFT exchange-correlation potential is only known approximately. As mentioned mentioned in section 3.1.4, most approximations to the exchange-correlation functional yield a seriously underestimated ionisation potential. While this leads to some interesting failures in the prediction of ground state properties, the situation is to some degree worse in TDDFT, as the number of unoccupied Kohn–Sham states with well-bound character is greatly reduced. Since an excitation in the LR-TDDFT formalism is represented through a linear combination of Kohn–Sham transitions from occupied to unoccupied states, this causes higher energy excitations that should have a well-bound character to contain significant contributions from unbound continuum Kohn–Sham states [85]. From a theoretical point of view, this means that the excitations have the wrong character, while from a practical point of view, very delocalised continuum Kohn–Sham states might not be well-represented depending on the basis set representation chosen in the ground state DFT calculation. These points are to be kept in mind when interpreting excitations with energies higher than the Kohn–Sham ionisation energy.

While the previous point is purely an effect of the failure of common ground state DFT approximations, there is a second class of errors in TDDFT excitation energies that originate from the approximation of f_{xc}. The first issue to be addressed is the locality approximation employed when using a (semi)-local exchange correlation functional in TDDFT. Following Peach *et al.* [86], one can define a measure spatial overlap between

the conduction Kohn–Sham state $\psi_c^{\mathrm{KS}}(\mathbf{r})$ and a valence Kohn–Sham state $\psi_v^{\mathrm{KS}}(\mathbf{r})$ in a system as

$$O_{cv} = \int \mathrm{d}^3 r \, |\psi_c^{\mathrm{KS}}(\mathbf{r})||\psi_v^{\mathrm{KS}}(\mathbf{r})|. \tag{4.60}$$

Then for a given excitation described by the elements x_{cv} of the transition eigenvector $\mathbf{x}$, one can define a measure Λ for the spatial overlap of the electron and the hole state, with

$$\Lambda = \frac{\sum_{cv} |x_{cv}|^2 O_{cv}}{\sum_{cv} |x_{cv}|^2}. \tag{4.61}$$

Now, consider the case where $\Lambda \to 0$, which corresponds to a complete spatial separation of the electron and the hole density associated with the excitation described by transition vector $\mathbf{x}$. In this case, the excitation associated with $\mathbf{x}$ is a charge-transfer excitation and it is trivial to show that $\mathbf{x}^\dagger \mathbf{K} \mathbf{x} \to 0$ for any (semi)-local exchange correlation kernel. Thus for a charge-transfer excitation, (semi)-local functionals predict an excitation energy that approaches that of pure Kohn–Sham eigenvalue differences and the TDDFT correction vanishes, which causes charge-transfer excitations to be seriously underestimated in energy. The same is true for infinite periodic systems, where the electron and the hole density completely delocalise, which has led to the statement that common (semi)-local TDDFT functionals fail for solids [77]. This problem can be addressed by making use of a hybrid functional, adding a fraction of exact Hartree–Fock exchange to the Tamm–Dancoff coupling matrix, such that

$$K^{\mathrm{HF}}_{cv,c'v'} = -c_x \int \mathrm{d}^3 r \, \mathrm{d}^3 r' \, \frac{\psi_{c'}^{\mathrm{KS}*}(\mathbf{r}) \psi_c^{\mathrm{KS}}(\mathbf{r}) \psi_{v'}^{\mathrm{KS}}(\mathbf{r}') \psi_v^{\mathrm{KS}*}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{4.62}$$

where c_x is the fraction of exact exchange used. Note that $K^{\mathrm{HF}}_{cv,c'v'}$ does add an explicit non-local coupling between the effective electron and hole wavefunction of the excitation and thus does not vanish, even if there is no spatial overlap between the electron and hole density. Introducing an exact-exchange term of the form of equation (4.62) yields a significant improvement in the description of some charge transfer excitations and excitations in extended systems. Furthermore, since there is now an explicit coupling between electron and hole densities in the system, $K^{\mathrm{HF}}_{cv,c'v'}$ can yield excitons with energies lower than the Kohn–Sham bandgap of the system [77].

While non-locality is clearly important for the description of long-range excitations, the effects of the frequency dependence of f_{xc} on the spectrum of excitation energies are less well understood [85]. However, if the ω-dependence of $\mathbf{H}^{[2p]}(\omega)$ is ignored, the TDDFT eigenvalue equation has the same dimensions as the number of possible Kohn–Sham transitions in the system, meaning that it can only yield excitations that can be described by a single electron-hole pair. If a ω-dependence is included, then the matrix

equation can have additional solutions with multi-electron-hole character [85]. However, in routine calculations on realistic systems, the frequency dependence in f_{xc} is generally ignored.

4.3 TDDFT and many-body perturbation theory

TDDFT has gained widespread popularity over recent years as the method of choice to calculate excited states of systems that are too large for more advanced quantum chemistry methods, both because of its conceptual simplicity and low computational cost and the fact that it often produces excitation energies that are of better quality than those obtained from the time-dependent Hartree–Fock method for states without charge transfer character. However, as seen in the previous section, there are problems with the method which mainly originate from the ambiguity in how to approximate the TDDFT kernel in different systems. Another popular approach in obtaining excitation spectra of solids and molecules is derived from many-body perturbation theory (MBPT). It has the advantage of providing more insight into what approximations are made in the calculation of excitation energies as compared to the relatively uncontrolled approximations made when choosing a functional form for f_{xc}.

For the purpose of this work these methods are not discussed in too much details as there are comprehensive reviews available (see for example Onida *et al.* [77]). However, a short discussion of the merits of MBPT from the perspective of the TDDFT approach is enlightening in terms of understanding the role of the TDDFT kernel as well as in justifying why TDDFT forms the only viable method currently available for obtaining excited states of the very large systems of interest in this work. While the basic theory of MBPT is entirely separate from Kohn–Sham DFT, here it will only be considered as a perturbative treatment to improve a Kohn–Sham starting point obtained from a DFT Hamiltonian with some approximation to the exchange-correlation potential.

4.3.1 Green's function methods

While in the density functional methods discussed previously in this work, the main quantity of interest is the single-particle electron density or the single-particle density matrix, in MBPT the quantity of interest is the single-particle Green's function. Following Sadovskii [87], the Green's function method is introduced by considering a single particle, time-independent Hamiltonian $\hat{H}_{\mathrm{sp}}$ such that

$$\mathrm{i}\frac{\partial\psi(\mathbf{r},t)}{\partial t} - \hat{H}_{\mathrm{sp}}\psi(\mathbf{r},t) = 0 \tag{4.63}$$

Instead of solving the above equation for $\psi(\mathbf{r}, t)$, one can solve the quasi-Schrödinger equation for the single-particle Green's function $G(\mathbf{r}, t, \mathbf{r}', t')$, that is given by

$$\mathrm{i}\frac{\partial G}{\partial t} - \hat{H}_{\mathrm{sp}} G = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t') \tag{4.64}$$

subject to the initial condition $G(\mathbf{r}, t^+, \mathbf{r}', t) = \delta(\mathbf{r} - \mathbf{r}')$. Physically, the single-particle Green's function represents the probability amplitude of a particle that is at position $\mathbf{r}'$ at time t' transitioning to position $\mathbf{r}$ at time t. The Green's function can then be used to express $\psi(\mathbf{r}, t + \tau)$ through $\psi(\mathbf{r}', t)$:

$$\psi(\mathbf{r}, t + \tau) = \int \mathrm{d}^3 r' \, G(\mathbf{r}, t + \tau, \mathbf{r}', t)\psi(\mathbf{r}', t) \tag{4.65}$$

A simple substitution shows that the above expression obeys the time-dependent Schrödinger equation as long as one assumes $G = 0$ for $\tau < 0$ in order to satisfy causality. Expanding the Green's function in terms of the eigenstates $\phi_i(\mathbf{r})$ of the single particle Hamiltonian, and noticing that since $\hat{H}_{\mathrm{sp}}$ is time-independent, the time-evolution of $\psi(\mathbf{r}, t)$ is easily described (see section 2.3), it follows that:

$$G_{ij}(\tau) = \int \mathrm{d}^3 r \, \mathrm{d}^3 r' \, \phi_i(\mathbf{r}) G(\mathbf{r}, t + \tau, \mathbf{r}', t)\phi_j(\mathbf{r}') = \delta_{ij} e^{-\mathrm{i}\epsilon_j \tau}\theta(\tau) \tag{4.66}$$

where $\theta(\tau)$ is the Heaviside step function defined as $\theta = 1$ for $\tau \geq 0$ and $\theta = 0$ otherwise. Thus, G can be written as a function of the time difference τ only and can therefore, under a simple Fourier transform, be expressed in energy space:

$$G_{jj}(\omega) = \frac{1}{\omega - \epsilon_j + \mathrm{i}\delta}. \tag{4.67}$$

Here, δ is an infinitesimal positive number ensuring that the inverse Fourier transform of $G(\omega)$ into the time domain yields equation (4.66) [21].

Now consider a many-electron system described by a non-interacting Hamiltonian. In principle, this can be any noninteracting time-independent Hamiltonian, but for the purpose of this work, the Hamiltonian is considered to be describing the fictitious Kohn–Sham system introduced in 3.1.3, where $\{\psi_i^{\mathrm{KS}}\}$ and $\{\epsilon_i^{\mathrm{KS}}\}$ define the Kohn–Sham eigenstates and eigenvalues of the system. Considering that all states below the Fermi-level ϵ_f are occupied and all states above are unoccupied, it is possible to write down Green's functions $G_{ij}^+(\tau)$ describing the evolution of occupied eigenstates and $G_{ij}^-(\tau)$ describing the evolu-

[21] In order to perform the inverse Fourier transform, one has to make use of Cauchy's theorem and consider $G(\omega)$ as a function of the complex variable ω in the lower half of the complex plane. The contour chosen for the integration is a semicircle in the lower half plane, where the factor $\mathrm{i}\delta$ ensures that the pole is inside the contour region.

tion of unoccupied eigenstates. It is then convenient to define the total Green's function $G_{ij}(\tau)$ so that it describes the evolution of electrons for $\tau > 0$ and the evolution of holes for $\tau < 0$. This can be done by defining

$$G_{ij}(\tau) = \begin{cases} G_{ij}^{+}(\tau) & \tau > 0 \\ -G_{ij}^{-}(-\tau) & \tau < 0 \end{cases} \tag{4.68}$$

yielding a total expression for $G(\mathbf{r}, \mathbf{r}', \omega)$ given by

$$G(\mathbf{r}, \mathbf{r}', \omega) = \sum_j \frac{\psi_j^{\mathrm{KS}}(\mathbf{r})\psi_j^{\mathrm{KS}}(\mathbf{r}')}{\omega - \epsilon_j^{\mathrm{KS}} + \mathrm{i}\delta\mathrm{sign}(\epsilon_f - \epsilon_j^{\mathrm{KS}})} \tag{4.69}$$

where the sign function is introduced to ensure that if one considers the analytic continuation of $G(\omega)$ to the entire complex plane, the poles corresponding to unoccupied Kohn–Sham eigenstates are in the upper half plane while the poles corresponding to occupied Kohn–Sham states are in the lower half plane.

For the remainder of this work $G_0(\mathbf{r}, \mathbf{r}', \omega)$ shall refer to the Green's function of a noninteracting system defined through equation (4.69). Note that G_0 contains all information of the unoccupied and occupied eigenstates of the fictitious Kohn–Sham system. Let $G(\mathbf{r}, \mathbf{r}', \omega)$ then denote the fully interacting Green's function of the system, where the interacting potential is defined as the Coulomb potential v between individual electrons, minus contributions already accounted for in the mean-field treatment of the Kohn–Sham system.

If the interacting system can be described in terms of weakly interacting, well defined quasiparticles [88], the fully interacting Green's function has poles at complex quasiparticle energies $\{\epsilon_j^{\mathrm{QP}}\}$, where $\mathrm{Re}\left[\epsilon_j^{\mathrm{QP}}\right]$ corresponds to the electron addition or removal energy of a single electron of the N-electron system such that

$$\mathrm{Re}\left[\epsilon_j^{\mathrm{QP}\{+\}}\right] = E_j(N+1) - E_0(N) \tag{4.70}$$

$$\mathrm{Re}\left[\epsilon_j^{\mathrm{QP}\{-\}}\right] = E_j(N-1) - E_0(N) \tag{4.71}$$

and the imaginary part $\mathrm{Im}\left[\epsilon_j^{\mathrm{QP}}\right]$ signifies that the quasiparticle state has a finite lifetime due to scattering events with other quasiparticle states. In the limit of strong interactions, G ceases to have well defined quasiparticle poles and describing the system in any form of single particle-like picture becomes an ill-defined concept [88].

It is then possible to derive the fully interacting Green's function from the noninteracting Green's function through a Dyson equation. To do so, one imagines that a quasiparticle propagates freely for some time in the system, until some interaction or

scattering event happens, after which it propagates freely again. Defining the self-energy Σ as the sum of all irreducible scattering processes of the single particle, and summing over all possible numbers irreducible scattering events a particle can undergo, one obtains the symbolic equation

$$G = G_0 + G_0\Sigma G_0 + G_0\Sigma G_0\Sigma G_0 + \cdots \tag{4.72}$$

which can be written as [87]

$$G(1,2) = G_0(1,2) + \int \mathrm{d}\tau_3\, \mathrm{d}\tau_4 G_0(1,3)\Sigma(3,4)G(4,2) \tag{4.73}$$

where the short hand notation of $1 = \mathbf{r}_1, t_1$ has been introduced. Thus, $\Sigma(1,2)$ represents a compact way to express all possible changes of a particle's motion as a reaction to interactions with other particles in the system. Solving for G through (4.73) then yields the entire quasiparticle spectrum of the system, containing the lifetimes and energies corresponding to all single electron additions and removals to the N-electron interacting system. Thus G contains information on all single-particle excitations of the system and can furthermore be used to express the ground state expectation value of any single-particle operator [87].

4.3.2 The GW-approximation and the Bethe-Salpeter equation

While $\Sigma(1,2)$ contains all information needed to solve the many-electron system starting from an arbitrary mean-field Hamiltonian, it cannot be calculated in practice as this evaluation would correspond to a full solution of the many-body problem. In general, Σ is only known in terms of an infinite perturbation expansion and has to be approximated in actual calculations. A practical approach to such an approximation is given by Hedin [89], who expanded the self-energy in terms of a dynamically screened Coulomb interaction $W(\mathbf{r}, \mathbf{r}', \omega)$ rather than the bare particle-particle interaction $v(\mathbf{r}, \mathbf{r}')$. From a physical point of view, $W(1,2)$ represents the potential at point 1 due to a test charge at 2, where all effects due to the polarisation of the many-electron system are included, yielding a screening of the interaction [89]. Hedin derived a set of five coupled equations that, when solved self-consistently, yield the full interacting Green's function of the system. However, in most practical calculations, only the simplifying approximation to Hedin's equations is considered, which is known as the GW approximation to the self-energy Σ.

The GW-approximation can be written in compact form as

$$\begin{aligned} P(1,2) &= -\mathrm{i}G(1,2)G(2,1) && (4.74)\\ W(1,2) &= v(1,2) + \int \mathrm{d}\tau_3\,\mathrm{d}\tau_4\, v(1,3)P(3,4)W(4,2) \\ &= \int \mathrm{d}\tau_3\, v(1,3)\epsilon^{-1}(3,2) && (4.75)\\ \Sigma(1,2) &= \mathrm{i}\int \mathrm{d}\tau_3\, G(1,3)W(3,2) && (4.76)\\ G_{\mathrm{new}}(1,2) &= G(1,2) + \int \mathrm{d}\tau_3\,\mathrm{d}\tau_4 G(1,3)\Sigma(3,4)G_{\mathrm{new}}(4,2) && (4.77) \end{aligned}$$

Here, $\epsilon^{-1}(1,2)$ is the inverse dielectric function of the system and $v(1,2) = \delta(t_1 - t_2)$ $|\mathbf{r}_1 - \mathbf{r}_2|^{-1}$. The four equations above have to be solved self-consistently with $G = G_0$ as a starting point. However, since the above set of equations is only an approximation to the full self-consistent Hedin's equations, it is not guaranteed that full self-consistency delivers better results than a simple non-selfconsistent perturbative correction that is obtained when only applying the equations once starting from some G_0. For this reason, GW is most commonly performed in the so-called G_0W_0 approximation, which corresponds to a one-shot perturbative correction to a Kohn–Sham DFT starting point [90].

Once the above system of equations has been solved and some approximation of $\Sigma(1,2)$ has been obtained, it is then possible to solve for the quasiparticle eigenstates and eigenenergies through

$$\hat{H}_{KS}\psi_j^{\mathrm{QP}}(\mathbf{r}) + \int \mathrm{d}^3r' \left[\Sigma(\mathbf{r},\mathbf{r}',\epsilon_j^{\mathrm{QP}}) - \delta(\mathbf{r}-\mathbf{r}')V_{\mathrm{xc}}(\mathbf{r})\right]\psi_j^{\mathrm{QP}}(\mathbf{r}') = \epsilon_j^{\mathrm{QP}}\psi_j^{\mathrm{QP}}(\mathbf{r}). \quad (4.78)$$

Note that the above equation is non-linear due to the energy-dependence of the self-energy. Assuming that the eigenenergies and eigenstates of the Kohn–Sham system form a good approximation to the quasiparticle energies and eigenstates, a first order correction to the Kohn–Sham energies can be found as [77, 90]

$$\epsilon_j^{\mathrm{QP}} = \epsilon_j^{\mathrm{KS}} + Z_j\langle\psi_j^{\mathrm{KS}}|\Sigma(\epsilon_j^{\mathrm{KS}}) - V_{\mathrm{xc}}|\psi_j^{\mathrm{KS}}\rangle \quad (4.79)$$

with

$$Z_j^{-1} = 1 - \left\langle \psi_j^{\mathrm{KS}} \left| \frac{\partial\Sigma(\omega)}{\partial\omega}\right|_{\epsilon_j^{\mathrm{KS}}} \right| \psi_j^{\mathrm{KS}} \right\rangle. \quad (4.80)$$

Thus the GW-approximation allows for the calculation of electron and hole quasiparticle energies starting from a Kohn–Sham Hamiltonian. In general, the quasiparticle eigenstates obtained yield a significant improvement over the Kohn–Sham bandgap for a wide range of semiconductors [77]. However, for the purpose of this work, quantities

of interest are not the electron addition and removal energies, but rather in neutral excitations to the system. Thus it is necessary to consider an interaction between electron and hole Green's functions. Introducing the non-interacting two particle electron-hole Green's function $L_0(1,2,1',2') = G(1,2')G(2,1')$ which contains all single particle correlation effects due to the presence of the interacting single particle Green's function, the Bethe-Salpeter equation for the fully interacting electron-hole Green's function can be written as [84]:

$$L(1,2,1',2') = L_0(1,2,1',2') + \int \mathrm{d}\tau_3\,\mathrm{d}\tau_4\,\mathrm{d}\tau_5\,\mathrm{d}\tau_6\, L_0(1,4,1',4)K(3,5,4,6)L(6,2,5,2') \tag{4.81}$$

Here, $K(3,5,4,6)$ is the interaction kernel that contains all possible irreducible interactions between the electron and the hole Green's functions. If one considers only spontaneous generation and annihilation of electron-hole pairs, only two of the four free time variables in the above equation are independent and thus the entire time dependence can again be expressed through a single variable τ of the time difference between the two independent time variables [84]. It is then possible to transform the entire equation into energy space with a single Fourier transform.

If one expands the entire spatial dependence of equation (4.81) in terms of quasiparticle electron and hole wavefunctions $\{\psi_c^{\mathrm{QP}}\}$ and $\{\psi_v^{\mathrm{QP}}\}$, it is possible to obtain a similar eigenvalue equation of an effective 2-particle Hamiltonian as in the TDDFT case (see 4.2.4), which in the Tamm–Dancoff approximation can be expressed as

$$\sum_{c'v'} \left(H_{\mathrm{BSE}}^{[2p]}\right)_{cv,c'v'} A_{c'v'} = (\epsilon_c^{\mathrm{QP}} - \epsilon_v^{\mathrm{QP}})A_{cv} + 2\sum_{c'v'} K_{cv,c'v'}^{\mathrm{BSE}}(\Omega)A_{c'v'} = \Omega A_{cv} \tag{4.82}$$

where

$$\begin{aligned} K_{cv,c'v'}^{\mathrm{BSE}}(\Omega) \;=\; & \mathrm{i}\int \mathrm{d}^3r_3\,\mathrm{d}^3r_4\,\mathrm{d}^3r_5\,\mathrm{d}^3r_6\,\psi_v^{\mathrm{QP}}(\mathbf{r}_4)\psi_c^{\mathrm{QP}*}(\mathbf{r}_3)K(\mathbf{r}_3,\mathbf{r}_5,\mathbf{r}_4,\mathbf{r}_6;\Omega) \\ & \times\psi_{v'}^{\mathrm{QP}*}(\mathbf{r}_6)\psi_{c'}^{\mathrm{QP}}(\mathbf{r}_5) \end{aligned} \tag{4.83}$$

and in the GW-approximation, the kernel $K(3,5,4,6)$ is given by [84]

$$K(3,5,4,6) = -\mathrm{i}\delta(3,4)\delta(5^-,6)v(3,6) + \mathrm{i}\delta(3,6)\delta(4,5)W(3^+,4). \tag{4.84}$$

Here, it is again necessary to introduce positive and negative infinitesimal numbers to the time variables in order to make the Fourier transforms into and from energy space well defined.

Thus, in order to solve for the Bethe-Salpeter equation it is first necessary to per-

form a GW-calculation to some level of approximation, obtain all quasiparticle energies and quasiparticle wavefunctions from equation (4.78) and then build the Bethe-Salpeter eigenvalue equation using the screened Coulomb interaction W. One then can solve the nonlinear eigenvalue equation for the true electron-hole excitation energies of the system.

In most cases, the Bethe-Salpeter equation in the Tamm-Dancoff approximation is actually solved in an approximate fashion. First, the frequency dependence of $W(\mathbf{r}, \mathbf{r}', \omega)$ is often ignored, turning the nonlinear eigenproblem into a conventional Hermitian eigenproblem that is much simpler to solve. Secondly, one often makes the approximation that $\psi_j^{\mathrm{QP}}(\mathbf{r}) \approx \psi_j^{\mathrm{KS}}(\mathbf{r})$, ie. one assumes that the Kohn–Sham wavefunctions for the electron and the hole state form a good approximation to the quasiparticle wavefunctions. In that case, equation (4.82) can be expanded in terms of Kohn–Sham eigenstates and only the first order correction to the Kohn–Sham eigenvalues from equation (4.79) is used in the diagonal part of the effective Bethe-Salpeter Hamiltonian.

The Bethe-Salpeter equation has been shown to deliver very good agreements with experimental data, especially for infinite systems, where TDDFT kernels struggle to correctly reproduce excitonic effects [77].

4.3.3 The TDDFT exchange-correlation kernel from a MBPT perspective

Superficially, the eigenvalue equation of the effective 2-particle Hamiltonian of the Bethe-Salpeter equation (equation (4.82)) has a similar form to that of the TDDFT eigenvalue equation in the linear response formalism derived in section 4.2.4. Interestingly, while the interaction kernel in TDDFT is dependent on the exchange-correlation kernel, a nonlocal, frequency-dependent function that has an unknown functional form of the density, the interaction kernel in the Bethe-Salpeter equation is dependent on $W(\mathbf{r}, \mathbf{r}', \omega)$, which has a very straightforward physical interpretation as a dynamically screened Coulomb interaction.

The main difference between the two effective 2-particle Hamiltonian equations is that the TDDFT equation is derived from a two-point quantity, the linear density response $\chi(\mathbf{r}, \mathbf{r}', \omega)$, while the Bethe-Salpeter eigenvalue equation is derived from the 4-point electron-hole correlation function $L(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_1', \mathbf{r}_2', \omega)$. This causes the interaction kernel $K(\Omega)$ of the Bethe-Salpeter equation to become a 4-point quantity, as compared to the corresponding 2-point quantity in TDDFT [77]. To see the difference in detail, consider only the coupling part of the 2-particle Hamiltonian, which is assumed to be frequency-independent, and for simplicity it is assumed that the electron and hole quasiparticle wavefunctions are equal to the Kohn–Sham electron and hole wavefunctions. The system at hand is also assumed to be isolated so that the Kohn–Sham eigenfunctions can be chosen to be real. Let $\mathbf{x}$ then denote a trial vector corresponding to some linear

combination of Kohn–Sham electron-hole states. Furthermore, the electron-hole response density and response density matrix associated with the trial vector $\mathbf{x}$ can be defined as

$$\rho^{\{1\}}(\mathbf{r}) = \sum_{cv} \psi_c^{\mathrm{KS}}(\mathbf{r}) x_{cv} \psi_v^{\mathrm{KS}}(\mathbf{r}) \tag{4.85}$$

$$\rho^{\{1\}}(\mathbf{r}, \mathbf{r}') = \sum_{cv} \psi_c^{\mathrm{KS}}(\mathbf{r}) x_{cv} \psi_v^{\mathrm{KS}}(\mathbf{r}'). \tag{4.86}$$

Then the interaction energy due to the kernel K for a trial vector $\mathbf{x}$ for both the Bethe-Salpeter equation and the TDDFT eigenvalue equation can be written as

$$\mathbf{x}^\dagger \mathbf{K}_{\mathrm{TDDFT}} \mathbf{x} = 2 \int \mathrm{d}^3 r \, \mathrm{d}^3 r' \left[\frac{\rho^{\{1\}}(\mathbf{r}) \rho^{\{1\}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \rho^{\{1\}}(\mathbf{r}) \left. \frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}')} \right|_{\rho_0} \rho^{\{1\}}(\mathbf{r}') \right] \tag{4.87}$$

$$\mathbf{x}^\dagger \mathbf{K}_{\mathrm{BSE}} \mathbf{x} = 2 \int \mathrm{d}^3 r \, \mathrm{d}^3 r' \left[\frac{\rho^{\{1\}}(\mathbf{r}) \rho^{\{1\}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \rho^{\{1\}}(\mathbf{r}, \mathbf{r}') W(\mathbf{r}, \mathbf{r}') \rho^{\{1\}}(\mathbf{r}', \mathbf{r}) \right] \tag{4.88}$$

As it can be seen from the two equations above, the first term in the interaction kernel is identical in both cases and corresponds to the Hartree energy of the electron-hole density. The second term however has a different structure due to the δ-functions present in the Bethe-Salpeter kernel. While the exchange-correlation part of the TDDFT kernel still acts on Kohn–Sham transition densities, in the Bethe-Salpeter equation equivalent the screened Coulomb interaction W acts explicitly on electron-hole density matrices, thus introducing a long-range coupling of electron and hole wavefunctions. Therefore W trivially describes long-range electron-hole interactions in charge-transfer states or delocalised systems, while in any (semi)-local approximation to the exchange-correlation potential, this interaction is necessarily missing in TDDFT.

In the linear-response TDDFT method, hybrid functionals have been suggested as a solution to the problem of missing long-range electron-hole interactions. It can be seen why this suggestion is sensible from a MBPT perspective by remembering that W is just an infinite series summation in terms of the bare Coulomb interaction v, thus substituting v for W would lead to the second term in equation (4.88) to be simply equal to the Hartree–Fock energy associated with the electron-hole density matrix. Thus in the long-range charge-transfer limit, where the (semi)-local part of any hybrid functional vanishes, the electron-hole interaction W in hybrid-TDDFT is simply approximated by a fraction c_x of the bare Coulomb interaction. Given that W is related to v through the inverse dielectric function such that $W = v\epsilon^{-1}$ in symbolic notation (see equation (4.75)), using a hybrid functional in the long-range limit becomes equivalent to approximating the BSE interaction W by assuming a static, constant dielectric function of $\epsilon = 1/c_x$. This can be seen as a non-rigorous justification for expecting hybrid functionals in TDDFT

to deliver relatively good results in uniform systems, where the spatial structure of W is negligible. Whether many-body effects can be resolved in more complicated situations of non-uniform screening, like in a graphene layer deposited on a substrate [91], is more doubtful.

The close similarity between the structures of the two-particle Hamiltonians in linear-response TDDFT and the Bethe-Salpeter equation have led to the derivation of an effective exchange-correlation kernel f_{xc} from the GW+BSE approximation capable of reproducing all long-range excitonic effects in bulk materials that are correctly accounted for in the BSE [92]. However, the kernel can only be constructed by first solving the GW-approximation and thus does not represent a significant computational saving compared to the full Bethe-Salpeter equation. In conclusion, it can be said that while TDDFT is formally exact for the exact exchange-correlation kernel, the similarity between the TDDFT and MBPT approach allows a more detailed analysis of the failures of certain approximations to the exchange-correlation kernel and helps in recognising the important effects that are ignored in simple approximations to f_{xc}.

4.4 Scaling considerations: TDDFT vs. MBPT

For the remainder of this chapter the computational scaling of the two main TDDFT approaches, the real-time and the linear-response method, are briefly considered and compared to that of the GW-method in connection with the Bethe-Salpeter equation. As can be appreciated from the previous discussion, the computational complexity of a TDDFT calculation is strongly dependent on the sophistication of the approximation to the exchange-correlation kernel f_{xc} used. Here the discussion shall be limited to standard (semi)-local and hybrid functionals for f_{xc} and thus ignore any GW+BSE derived kernels like the one discussed in [92], as they have a similar computational complexity as the full MBPT solution to the excited states of the system. Any memory effects in both the TDDFT and the MBPT approaches shall also be ignored.

While the full solution to the GW+BSE equation has the potential to provide a better description of excitonic effects than most commonly used approximations to the exchange-correlation kernel, the GW calculation can be seen as a major computational bottleneck. Generally, GW calculations do not show a scaling better than $\mathcal{O}(N^3)$ with the number of atoms in the system [93–96]. This is mainly due to the non-locality of all functions involved, as well as the fact that one has to perform integrals over the entire frequency range of functions that only decay very slowly (as $1/\omega$) with frequency. A naive construction of the two-particle Hamiltonian and a direct diagonalisation of the BSE scales as $\mathcal{O}(N^6)$, as the number of quasiparticle electron-hole pairs scales as $\mathcal{O}(N^2)$

with system size, yielding a total number of matrix elements that scale as $\mathcal{O}(N^4)$. Since in large systems, only a few low energy excited states age generally of interest rather than the entire spectrum, one can make use of iterative techniques that avoid the explicit construction and diagonalisation of $\mathbf{H}_{\mathrm{BSE}}^{[2p]}$. However, one still has to evaluate the result of a trial vector acting on $\mathbf{K}_{\mathrm{BSE}}$, which, due to the non-locality of $W(\mathbf{r}, \mathbf{r}', \omega)$, has a scaling of $\mathcal{O}(N^4)$ associated with it. It becomes clear that this unfavourable scaling makes the GW+BSE approach prohibitive for large systems that already rely on linear-scaling DFT approaches to solve for the ground state.

The scaling of TDDFT with most commonly used functionals however is considerably more favourable. It has been known for over a decade that the real-time TDDFT approach can be made to scale linearly with system size, if instead of individual Kohn–Sham orbitals, the single particle density matrix is propagated in time and its sparsity is exploited [97][22]. For the linear-response TDDFT approach, it has recently been noted that, when the solution of the eigenvalue equation to a small number of excited states of interest, these excitation energies can be obtained in linear-scaling effort using iterative eigensolvers [98–100]. Thus, TDDFT is the method of choice for the calculation of excited states of large systems, even though common exchange-correlation functional approximations cannot yield the full quality of excitonic effects that can be obtained from the solution of the Bethe-Salpeter equation.

Thus the remaining question is whether to use the time evolution approach to TDDFT or the linear-response formalism. Since for the purpose of this work, the focus lies on calculating only a small number of excitations in the low energy range of very large systems, the linear-response formalism is considered to be the most appropriate approach, since the time-evolution TDDFT approach generates the entire spectrum. Furthermore, the linear-response approach allows for the calculation of the full transition vector of any excitation, which contains additional information on the excited state potential energy surface, that are not available in the real-time TDDFT approach. The RT-TDDFT approach has the additional disadvantage that the Kohn–Sham equations need to be propagated for a rather long time to get a good resolution on the spectrum, which might be prohibitive for very large systems. Thus for the remainder of this work, TDDFT is considered in the linear-response formalism only.

[22] It should be noted that there is no full formal justification why the perturbed density matrix should show the same sparsity properties as the ground state density matrix. A more detailed discussion on the question of sparsity of response density matrices is delivered in section 6.2

Chapter 5

The ONETEP code

Following the discussion in the previous chapters, the aim of this work is to develop an extension of the linear-scaling DFT code ONETEP [101,102] to allow for the calculation of the TDDFT excitation energies in the linear response formalism. The DFT implementation has been demonstrated to provide an excellent parallel scaling, allowing for large scale calculations of ground state properties, while retaining the well-controlled convergence properties of plane-wave DFT approaches. The task is to develop a TDDFT extension to the DFT approach that retains all desirable features of the ground state method, while allowing for the calculation of low energy optical spectra of previously inaccessible system sizes. In order to derive a linear-scaling TDDFT approach, it is first necessary to briefly introduce some of the special features of the DFT implementation as compared with conventional linear-scaling approaches. First the approach to solving the ground state DFT problem in the ONETEP code is presented, before introducing some more advanced functionality that is made use of in later chapters.

5.1 Linear-scaling DFT in ONETEP

The focus in this section is on the main ingredients of a standard ground state DFT calculation in ONETEP. The code makes use of many of the general linear-scaling DFT techniques presented in section 3.3, like the density-matrix formalism and the LNV method. Here, only the unique features of the approach are focused on that are not generally shared with other linear-scaling DFT implementations.

5.1.1 Psinc basis set

The ONETEP code makes use of periodic boundary conditions as described in section 3.2.1. Since the method is aimed at large system sizes, it can be assumed that the real space unit cell is large, meaning that in reciprocal space the Brillouin zone sampling is well-represented by the Γ-point. From this it follows directly that all eigenfunctions of the

DFT Hamiltonian can be chosen to be real and any **k**-point dependence can be dropped from the calculation.

While plane waves are a natural basis set choice for a system in periodic boundary conditions, they are not advantageous for linear-scaling methods due to their delocalised character in real space. Instead, an appropriate basis set for linear scaling purposes can be defined through psinc functions [103, 104], also referred to as bandwidth limited Dirac delta functions. Consider a simulation cell with real-space and reciprocal space lattice vectors $\{\mathbf{a}_i\}$ and $\{\mathbf{b}_j\}$, which is divided into a discrete grid with $N_i = 2J_i + 1$ grid points in the direction i (here, J_i is an integer). The psinc basis functions are then defined as

$$D_{KLM}(\mathbf{r}) \equiv D(\mathbf{r} - \mathbf{r}_{KLM}) = \frac{1}{N_1 N_2 N_3} \sum_{p=-J_1}^{J_1} \sum_{q=-J_2}^{J_2} \sum_{s=-J_3}^{J_3} e^{\mathrm{i}(p\mathbf{b}_1 + q\mathbf{b}_2 + r\mathbf{b}_3)\cdot(\mathbf{r}-\mathbf{r}_{KLM})} \tag{5.1}$$

with p, q and s being integers and $\mathbf{r}_{KLM}$ given by

$$\mathbf{r}_{KLM} = \frac{K}{N_1}\mathbf{a}_1 + \frac{L}{N_2}\mathbf{a}_2 + \frac{M}{N_3}\mathbf{a}_3. \tag{5.2}$$

Here, K, L and M are again integers. Thus the basis function $D_{KLM}(\mathbf{r})$ is centered at grid point $\mathbf{r}_{KLM}$ and has the desirable property that it is zero at all other grid points. Furthermore, a unitary transformation relates the set $\{D_{KLM}(\mathbf{r})\}$ to the basis of plane-waves that can be represented by the real space grid of the simluation cell [103]. In other words, the basis set is equivalent to a plane wave basis and the basis set quality can, like in plane wave methods, be expressed in terms of a kinetic energy cutoff. The psinc basis thus allows a systematic convergence of the DFT energy with basis set size, just like the plane-wave basis in conventional DFT approaches.

Since the psinc function $D_{KLM}(\mathbf{r})$ is zero on all other grid points apart from $\mathbf{r}_{KLM}$, any function that is localised to some region in real space can be expanded by the subset of $\{D_{KLM}(\mathbf{r})\}$ corresponding to grid points localised in the region. It is to be noted that while each psinc function is zero on all grid points apart from the one it is centered on, it has nonzero values between grid points. Thus any function expanded in terms of a subset of psincs in some localisation region is delocalised through the entire simulation cell, but only has nonzero values on grid points inside the localisation region. However, it can be shown that the overlap integrals between any function with the periodicity of the simulation cell and a function that is represented by $\{D_{KLM}(\mathbf{r})\}$ can be evaluated exactly as a summation over grid points $\mathbf{r}_{KLM}$ [103], meaning localisation constraints can be exploited when performing integrals.

5.1.2 NGWF representation

In line with other linear-scaling DFT implementations, the approach considered here relies on expanding the single-particle density matrix in terms of localised functions (see equation (3.73)). In ONETEP, the expansion is achieved through so-called Non-Orthogonal Generalised Wannier functions (NGWFs) $\{\phi_\alpha\}$, that are taken to be atom-centered functions confined within a strict localisation radius r_c which is predefined for each atomic species in the calculation.

The NGWFs are expanded in the underlying basis of psinc functions $\{D_{KLM}(\mathbf{r})\}$, such that each NGWF is defined through a set of expansion coefficients $\{C_{KLM,\alpha}\}$, with [104]

$$\phi_\alpha(\mathbf{r}) = \sum_{K=0}^{N_1-1}\sum_{L=0}^{N_2-1}\sum_{M=0}^{N_3-1} C_{KLM,\alpha} D_{KLM}(\mathbf{r}). \tag{5.3}$$

Note that due to the strict localisation constraint on the NGWFs, the number of non-zero expansion coefficients $C_{KLM,\alpha}$ per NGWF stays constant with system size. Due to the relationship between psinc functions and plane waves, the NGWFs can equally be written in a plane wave expansion, such that

$$\phi_\alpha(\mathbf{r}) = \frac{1}{V}\sum_{p=-J_1}^{J_1}\sum_{q=-J_2}^{J_2}\sum_{s=-J_3}^{J_3} \tilde{\phi}_\alpha(p\mathbf{b}_1 + q\mathbf{b}_2 + r\mathbf{b}_3) e^{\mathrm{i}(p\mathbf{b}_1+q\mathbf{b}_2+r\mathbf{b}_3)\cdot\mathbf{r}} \tag{5.4}$$

where $\tilde{\phi}_\alpha(p\mathbf{b}_1 + q\mathbf{b}_2 + r\mathbf{b}_3)$ is obtained from a discrete Fourier transform of the expansion coefficients $C_{KLM,\alpha}$.

The fact that the NGWFs used to represent the single particle density matrix and the Kohn–Sham Hamiltonian are expressed through an underlying basis means that the density matrix representation is not required to stay fixed during the calculation. The DFT total energy E_0 can thus be seen as a functional of both the density matrix in a given NGWF representation and the NGWF expansion coefficients $\{C_{KLM,\alpha}\}$.

5.1.3 The FFT box technique

The evaluation of the DFT kinetic energy in the NGWF representation can be seen as an example outlining the advantages of the underlying psinc basis used. Here a spin-degenerate system is assumed, with a ground state density matrix $\mathbf{P}^{\{v\}}$ expanded through a set of NGWFs $\{\phi_\alpha\}$. The Kohn–Sham kinetic energy of such a system can be simply written as

$$T_s[\rho] = -P^{\{v\}\alpha\beta}\langle\phi_\beta|\nabla^2|\phi_\alpha\rangle \tag{5.5}$$

Here, the advantage of the underlying basis becomes apparent. While applying the Laplacian operator in other localised basis set representations that are commonly used can be cumbersome, its action on a plane wave basis in reciprocal space is trivially evaluated. Since the NGWFs can be transformed into a plane wave basis via equation (5.4), a way the above matrix elements can be evaluated is by performing a Fast-Fourier transform (FFT) on ϕ_α, applying the Laplacian operator and transforming back, before performing the integral with ϕ_β in real space as a summation over grid points [103]. However, since both number of grid points in the simulation cell as well as the number of NGWFs scale as $\mathcal{O}(N)$ with system size, performing an FFT for all NGWFs to evaluate the kinetic energy matrix has a scaling of $\mathcal{O}(N^2\log(N))$ with system size associated with it.

Linear scaling can only be achieved by exploiting the strict localisation of the NGWFs. Rather than performing the FFT on the entire cell, a smaller region called an FFT box that contains the NGWF in real space can be used. Since the size of the NGWF localisation region is system size-independent, the size of the FFT box is also independent of system size, meaning that Fast-Fourier transforms of NGWFs have a constant computational scaling with simulation cell size associated with them. There are a number of requirements that define the minimum possible size of the FFT box, which can be found in reference [105]. Here it is just noted that the FFT box has to be chosen large enough so that for a given NGWF it can contain that NGWF at its centre and all its overlapping neighbours. The size of the FFT box in comparison to that of the entire simulation cell is a good indicator for the efficiency of the calculation compared with conventional approaches, where the crossover point with standard $\mathcal{O}(N^3)$ plane-wave methods is only reached once the simulation cell is considerably larger than the FFT box.

5.1.4 The two-step energy minimisation

In order to perform an energy minimisation in `ONETEP` the DFT energy E is considered as an explicit functional of both the density matrix $\mathbf{P}^{\{v\}}$ and the NGWF expansion coefficients $\{C_{KLM,\alpha}\}$. It is thus possible to define a two-step nested minimisation procedure such that [104]

$$E_{\min} = \min_{\{C_{KLM,\alpha}\}} L\left[\{C_{KLM,\alpha}\}\right] \tag{5.6}$$

where

$$L\left[\{C_{KLM,\alpha}\}\right] = \min_{\mathbf{P}^{\{v\}}} E\left[\mathbf{P}^{\{v\}}, \{C_{KLM,\alpha}\}\right] \tag{5.7}$$

Here, the direct minimisation of the DFT energy in equation (5.7) is carried out for fixed NGWF expansion coefficients, taking into account the change of the Kohn-Sham potential $V_{\mathrm{KS}}(\mathbf{r})$ as a result of changes in $\mathbf{P}^{\{v\}}$ and thus $\rho(\mathbf{r})$, while the idempotency and normalisation constraints on $\mathbf{P}^{\{v\}}$ are enforced as described in section 3.3.5. The second

optimisation step with respect to the NGWF coefficients can then be performed by finding the gradient of L with respect to $C_{KLM,\alpha}$. The exact form of L differs depending on how the constraints of idempotency and normalisation are enforced. Here, only the gradient of the energy with respect to $C_{KLM,\alpha}$ is explicitly shown, ignoring extra terms from the idempotency and orthonormality constraints. The gradient can be written as [104]:

$$\frac{\partial L}{\partial C_{KLM,\alpha}} = \int \mathrm{d}^3 r \, \frac{\delta L}{\delta \phi_\alpha(\mathbf{r})} \frac{\partial \phi_\alpha(\mathbf{r})}{\partial C_{KLM,\alpha}} = \left[4w\hat{H}\phi_\beta(\mathbf{r}) P^{\{v\}\alpha\beta} \right]_{\mathbf{r}=\mathbf{r}_{KLM}} . \tag{5.8}$$

Here, w is the grid weight associated with each grid point. Since the above gradient is a contravariant quantity, it has to be transformed into a covariant quantity by multiplying with $\mathbf{S}^\phi$. It can then be used as a search direction in a conjugate gradient algorithm to improve the expansion coefficients of all NGWFs. Due to the localisation constraints on the NGWFs, as well as the use of the FFT box technique, the gradient of L with respect to all NGWF coefficients can be evaluated in linear scaling effort, provided $\mathbf{P}^{\{v\}}$ is sparse.

Thus once the two-step minimisation is converged, ONETEP yields a $\mathbf{P}^{\{v\}}$ in the ideal representation of NGWFs $\{\phi_\alpha\}$ for the given localisation constraint imposed.

5.1.5 Sparsity and linear scaling

As with other approaches, linear-scaling computational effort with system size can only be reached in ONETEP if the ground state density matrix can be truncated and becomes sparse in the representation of localised NGWFs $\{\phi_\alpha\}$. This is achieved by predefining a cutoff radius R_{cut} such that the matrix element $P^{\{v\}\alpha\beta} = 0$ for $|\mathbf{R}_\alpha - \mathbf{R}_\beta| > R_{\mathrm{cut}}$.

Together with the localisation radius r_c on the NGWFs, the cutoff radius on the density matrix elements predefines the sparsity pattern of any matrix encountered in a ground state calculation [106]. The atoms are then ordered in such a way that the dense matrix blocks of $\mathbf{P}^{\{v\}}$ are clustered around the diagonal. Both the exact knowledge of the sparsity pattern and the ordering of the atoms before the start of a calculation allows for optimised communication and efficient sharing matrix data over the CPUs available to the calculation [106].

In the LNV method (see section 3.3.5), the density matrix $\mathbf{P}^{\{v\}}$ is defined via the McWeeny purification transformation of an auxiliary density matrix $\mathbf{L}^{\{v\}}$, such that

$$\mathbf{P}^{\{v\}} = 3\mathbf{L}^{\{v\}}\mathbf{S}^\phi\mathbf{L}^{\{v\}} - 2\mathbf{L}^{\{v\}}\mathbf{S}^\phi\mathbf{L}^{\{v\}}\mathbf{S}^\phi\mathbf{L}^{\{v\}}. \tag{5.9}$$

Thus in order for there to be no truncation in the evaluation of the density matrix before the final result, the least sparse matrix that has to be explicitly evaluated is $\mathbf{L}^{\{v\}}\mathbf{S}^\phi\mathbf{L}^{\{v\}}$, which is considerably less sparse than $\mathbf{P}^{\{v\}}$. Only if the least sparse matrix required to be constructed in the calculation becomes sparse does the DFT calculation scale fully

linearly with system size.

However, it should be pointed out that the explicit optimisation of the NGWF expansion coefficients means that the NGWF set $\{\phi_\alpha\}$ is considerably smaller than comparable localised basis sets that stay fixed during the calculations. This has the effect that the system sizes at which point the dense matrix algebra begins to dominate calculation time are larger in ONETEP calculations than in methods using fixed localised orbitals and an explicit density matrix truncation is generally not necessary for system sizes up to about 2000 atoms.

5.1.6 Forces

One is often interested in nucleic forces in the context of density-functional theory calculations, as their evaluation is necessary to perform geometry optimisations to find equilibrium structures of systems of interest. To evaluate forces from a DFT ground state calculation, one has to consider the total derivative of the DFT energy with respect to the atomic positions $\mathbf{R}_\gamma$[23]. In the method described here, the NGWFs have an explicit dependence on the atomic positions, since they are all centered on specific atoms. Thus, the force acting on an individual atom γ can be written as

$$\mathbf{F}_\gamma = -\frac{\mathrm{d}E}{\mathrm{d}\mathbf{R}_\gamma} = -\frac{\partial E}{\partial \mathbf{R}_\gamma} - \frac{\partial E}{\partial P^{\{v\}\alpha\beta}}\frac{\partial P^{\{v\}\alpha\beta}}{\partial \mathbf{R}_\gamma} - \int \mathrm{d}^3 r \, \frac{\delta E}{\delta \phi_\alpha(\mathbf{r})}\frac{\partial \phi_\alpha(\mathbf{r})}{\partial \mathbf{R}_\gamma} \tag{5.10}$$

Since the density matrix is optimised at the end of a ground state calculation, the derivative of E with respect to $\mathbf{P}^{\{v\}}$ is zero and only two terms remain. One is the standard force term due to the partial derivative of the energy with respect to the atomic positions and the other one is a Pulay force term due to the derivative of the energy with respect to the NGWF representation. However, since a two-step minimisation is performed in ONETEP, the functional derivative of the energy with respect to $\{\phi_\alpha\}$ also vanishes at the end of a converged calculation and the force reduces to [107, 108]

$$\mathbf{F}_\gamma = -\frac{\partial E}{\partial \mathbf{R}_\gamma} \tag{5.11}$$

which is the familiar result obtained in plane-wave methods or other methods where the basis set used is independent of the atomic positions.

Since it is often not possible to converge the psinc expansion of the NGWFs to arbitrary

[23] In principle, forces can be computed via a finite difference approximation as well, but since this requires 6N individual ground state calculations (where N is the number of ions), the approach is impractical for the system sizes considered here.

accuracy[24], one is often left with a small gradient of the energy with respect to the NGWFs at the end of a ONETEP calculation. This can result in small but non-negligible Pulay force contributions, which can cause a slow convergence behaviour of the geometry optimisation algorithm. However, it is possible to derive a Pulay term correcting for the effect, which when included in the calculation, yields a significant speed up when performing geometry optimisations in many systems [108].

5.2 Relevant functionality

After having introduced the main features of the ONETEP code, the focus is now shifted to some of the more advanced functionalities that go beyond a simple evaluation of the ground state energy. Only functionality that will be used heavily in later chapters, either in technical derivations or in proof-of-principle calculations, is considered here.

5.2.1 Conduction NGWF optimisation

As seen in the previous section, the two-step energy minimisation strategy in ONETEP produces a valence density matrix $\mathbf{P}^{\{v\}}$ expanded in NGWFs $\{\phi_\alpha\}$. The psinc expansion coefficients of the individual NGWFs are optimised to form an ideal representation of $\mathbf{P}^{\{v\}}$. Thus, at the end of a converged ground state calculation, the DFT energy of the system is minimised with respect to variations in the localised NGWF representation. If the system has a well defined bandgap and the localisation radius of the NGWFs is chosen appropriately, the minimisation strategy guarantees to yield a representation of the occupied Kohn–Sham space that is comparable to that achievable by plane-wave methods.

However, in excited state methods, the interest is not only in a representation of the occupied subspace but also the unoccupied space, as it is necessary to represent electron-hole states. This poses a problem for the ONETEP approach, as the entire unoccupied space representable by a chosen NGWF set $\{\phi_\alpha\}$ can be written in form of a projection operator

$$\mathbf{P}^{\{c\}} = \left(\mathbf{S}^{\phi}\right)^{-1} - \mathbf{P}^{\{v\}}. \tag{5.12}$$

Here, $\mathbf{P}^{\{c\}}$ is to denote the effective density matrix representing the entire conduction space manifold in valence NGWF representation. While $\mathbf{P}^{\{v\}}$ is well represented by $\{\phi_\alpha\}$, the entire conduction space manifold is clearly only very poorly represented by $\mathbf{P}^{\{c\}}$ as the

[24]The reason why the minimisation with respect to the psinc expansion coefficients cannot be converged to the same degree as the minimisation with respect to the density matrix $\mathbf{P}^{\{v\}}$ is that there is an inherent conflict between the energy minimisation and the localisation constraint. Thus in practice, for well localised NGWFs, the gradient of the energy with respect to the NGWF expansion coefficients cannot be made arbitrarily small.

NGWF representation is far from complete and thus $|\phi_\alpha\rangle \left(S^\phi_{\alpha\beta}\right)^{-1} \langle\phi_\beta|$ is a poor representation of the identity operator. This has the effect that when one explicitly diagonalises $\mathbf{H}^\phi$ and compares the Kohn–Sham bandstructure energies to the ones generated by plane-wave codes, only the occupied bands are in good agreement while the agreement breaks down for the unoccupied space.

This shortcoming is not surprising. The size of the ONETEP NGWF representation is very small, even when compared to other localised orbital methods that do keep a fixed representation during the calculation. The *in situ* optimisation of the NGWFs necessarily creates an ideal representation for the occupied states, but there is no reason for the unoccupied states to be well represented. Given that the NGWF representation of $\mathbf{P}^{\{v\}}$ is justified via the existence of exponentially localised Wannier functions for the composite band of valence states, the representation of arbitrary parts of the conduction space manifold in the same manner is generally not valid.

However, it is reasonable to assume that a low energy subset of the conduction space can indeed be represented in a localised NGWF representation and that an explicit optimisation of the Kohn–Sham bandstructure energy with respect to the psinc expansion coefficients of the NGWF representation can yield results comparable to plane-wave methods. A technique to achieve this within the ONETEP code was recently introduced by Ratcliff and coworkers [109] and will be briefly outlined here.

Consider a single particle density matrix for a set of the lowest N_{virt} unoccupied Kohn–Sham states, such that, again assuming spin degeneracy of the ground state, one can write

$$\rho^{\{c\}}(\mathbf{r},\mathbf{r}') = \sum_{i=N_{\text{occ}}/2+1}^{(N_{\text{occ}}+N_{\text{virt}})/2} \psi_i^{\text{KS}}(\mathbf{r})\psi_i^{\text{KS}*}(\mathbf{r}'). \tag{5.13}$$

It is also assumed that the N_{virt} lowest unoccupied states form a set of composite bands (see section 3.2.2). In that case, $\rho^{\{c\}}(\mathbf{r},\mathbf{r}')$ can be rigorously expressed by some set of localised conduction NGWFs which will be referred to as $\{\chi_\alpha\}$ such that

$$\rho^{\{c\}}(\mathbf{r},\mathbf{r}') = \chi_\alpha(\mathbf{r})P^{\{c\}\alpha\beta}\chi_\beta(\mathbf{r}') \tag{5.14}$$

where $\mathbf{P}^{\{c\}}$ is now an effective density matrix in conduction NGWF representation spanning the lowest N_{virt} conduction states. The key of the conduction state optimisation in ONETEP is then to optimise both $\mathbf{P}^{\{c\}}$ and $\{\chi_\alpha\}$ to ideally represent the subspace of low energy conduction states. In order to do so a projected Kohn–Sham Hamiltonian in $\{\chi_\alpha\}$ representation is defined such that

$$\left(H^\chi_{\text{proj}}\right)^{\alpha\beta} = \langle\chi_\alpha|\hat{H}|\chi_\beta\rangle - \langle\chi_\alpha|\phi_\gamma\rangle \left(P^{\{v\}}H^\phi P^{\{v\}} - \lambda P^{\{v\}}S^\phi P^{\{v\}}\right)^{\gamma\delta} \langle\phi_\delta|\chi_\beta\rangle \tag{5.15}$$

where λ is a shift parameter. The effect of the projection in the Kohn–Sham Hamiltonian is to shift all valence states up in energy. If λ is chosen larger than the energy of the highest conduction state of interest, the valence states are shifted higher than that value and the lowest $N_{\text{virt}}/2$ eigenstates of the projected Hamiltonian correspond to the unoccupied Kohn–Sham bands of interest.

Thus, minimising the conduction bandstructure energy

$$E_{\text{BS}}^{\{c\}}\left[\mathbf{P}^{\{c\}}\right] = 2\text{Tr}\left[\mathbf{H}_{\text{proj}}^{\chi}\mathbf{P}^{\{c\}}\right] \tag{5.16}$$

with respect to $\mathbf{P}^{\{c\}}$, subject to the normalisation and idempotency constraints

$$\mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{P}^{\{c\}} = \mathbf{P}^{\{c\}} \tag{5.17}$$

$$2\text{Tr}\left[\mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\right] = N_{\text{virt}} \tag{5.18}$$

yields a density matrix in conduction NGWF representation spanning the lowest N_{virt} conduction states of the Kohn–Sham system. Note the similarity of the above expressions to normal ground state density-matrix DFT. However, unlike for the valence density matrix, there is no implicit self-consistency condition in the minimisation of equation (5.16), as the Kohn–Sham Hamiltonian is a functional of $\mathbf{P}^{\{v\}}$ and not $\mathbf{P}^{\{c\}}$. Thus, the optimisation of the conduction space can be carried out as a post processing step after the ground-state density matrix and the valence NGWFs of the system in question are calculated. The minimisation of the conduction bandstructure energy can then be carried out precisely in the same way as it is for the valence band structure energy, with the difference that the Kohn–Sham potential V_{KS} is kept fixed.

The usual two-step minimisation procedure is recovered by treating $E_{\text{BS}}^{\{c\}}$ as a functional of both $\mathbf{P}^{\{c\}}$ and $\{\chi_\alpha\}$. Following the same steps as in the ground-state density matrix optimisation procedure, the conduction optimisation thus yields a density matrix spanning a predefined set of low energy conduction states expressed by a set of NGWFs that is optimised to be an ideal representation for those states. Together, $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ then form an ideal representation for both the valence states and a low energy subset of the conduction space and Kohn–Sham energies produced from the two representations are in very good agreement with plane-wave results [109].

5.2.2 Linear-scaling PAW

In order facilitate the treatment of transition metal compounds, the PAW formalism (see section 3.2.7) is implemented in `ONETEP`. Due to the fact that the linear-scaling DFT approach used here is implemented in a density matrix formalism, a number of changes

have to be made to Blöchl's original formalism, which will be discussed here briefly.

Since the linear-scaling DFT formalism works with the single-particle density matrix rather than the individual Kohn–Sham states, the PAW transformation discussed in section 3.2.7 needs to be applied to the density matrix. Again denoting AE partial waves as $\{|\varphi_i\rangle\}$ and soft pseudo (PS) partial waves as $\{|\tilde{\varphi}_i\rangle\}$, the AE single-particle density matrix in operator form can be expressed in terms of the soft PS density matrix $\mathbf{P}^{\{v\}}$ such that

$$\rho^{\{v\}} = |\phi_\alpha\rangle P^{\{v\}\alpha\beta}\langle\phi_\beta| + \sum_{ij}\{|\varphi_i\rangle\langle\varphi_j| - |\tilde{\varphi}_i\rangle\langle\tilde{\varphi}_j|\}\,\langle\tilde{p}^i|\phi_\alpha\rangle P^{\{v\}\alpha\beta}\langle\phi_\beta|\tilde{p}^j\rangle \tag{5.19}$$

The matrix elements $(Q^i_\alpha)^\dagger = \langle\tilde{p}^i|\phi_\alpha\rangle$ and $Q^j_\beta = \langle\phi_\beta|\tilde{p}^j\rangle$ are effective projector-NGWF overlap matrices, projecting the NGWFs that are defined on the soft PS grid in the entire simulation cell onto the partial waves only defined within the augmentation spheres around the atoms. Using the projector, it is possible to define the effective valence density matrix inside the spheres as

$$\rho^{\{v\}ij} = \left(Q^i_\alpha\right)^\dagger P^{\{v\}\alpha\beta} Q^j_\beta. \tag{5.20}$$

In the PAW formalism, the AE density matrix rather than the soft PS density matrix has to obey the normalisation and idempotency constraint, such that it is required that

$$\rho^{\{v\}}\rho^{\{v\}} = \rho^{\{v\}} \tag{5.21}$$

$$2\mathrm{Tr}\left[\rho^{\{v\}}\right] = N_{\mathrm{occ}}. \tag{5.22}$$

In close analogy to the conventional PAW implementations outlined in section 3.2.7, the two constraints can be written in terms of the soft PS matrix $\mathbf{P}^{\{v\}}$ by simply using the effective linear PAW transformation operator τ and applying it to the NGWF overlap matrix. Thus, $\mathbf{S}^\phi$ is redefined such that

$$S^\phi_{\alpha\beta} = \langle\phi_\alpha|\tau^\dagger\tau|\phi_\beta\rangle = \langle\phi_\alpha|\left(1 + \sum_{ij}|\tilde{p}^i\rangle(\langle\varphi_i|\varphi_j\rangle - \langle\tilde{\varphi}_i|\tilde{\varphi}_j\rangle)\langle\tilde{p}^j|\right)|\phi_\beta\rangle. \tag{5.23}$$

Under this redefinition, the normalisation and idempotency constraints of the full AE density matrix can be expressed through the soft part $\mathbf{P}^{\{v\}}$ of the density matrix, with

$$\mathbf{P}^{\{v\}}\mathbf{S}^\phi\mathbf{P}^{\{v\}} = \mathbf{P}^{\{v\}} \tag{5.24}$$

$$2\mathrm{Tr}\left[\mathbf{P}^{\{v\}}\mathbf{S}^\phi\right] = N_{\mathrm{occ}} \tag{5.25}$$

Note that the above form of the normalisation and idempotency constraint is exactly the same as in the case of density matrix DFT using norm-conserving pseudopotentials and

all the AE effects due to the augmentation spheres are hidden within the redefinition of $\mathbf{S}^{\phi}$.

Now consider the full AE density $\rho(\mathbf{r})$. Following the definition of the density matrix, it becomes clear that the total AE density can be partitioned into three parts, one defined on the conventional real space grid and two others defined within the augmentation spheres around the atoms:

$$\rho(\mathbf{r}) = \tilde{\rho}(\mathbf{r}) - \tilde{\rho}^{(1)}(\mathbf{r}) + \rho^{(1)}(\mathbf{r}) \tag{5.26}$$

where

$$\tilde{\rho}(\mathbf{r}) = \phi_{\alpha}(\mathbf{r}) P^{\{v\}\alpha\beta} \phi_{\beta}(\mathbf{r}) \tag{5.27}$$

$$\rho^{(1)}(\mathbf{r}) = \sum_{ij} \varphi_i(\mathbf{r}) \rho^{\{v\}ij} \varphi_j(\mathbf{r}) \tag{5.28}$$

$$\tilde{\rho}^{(1)}(\mathbf{r}) = \sum_{ij} \tilde{\varphi}_i(\mathbf{r}) \rho^{\{v\}ij} \tilde{\varphi}_j(\mathbf{r}). \tag{5.29}$$

Thus, the total density is constructed by taking the soft PS density on the entire grid, subtracting the soft part within the augmentation spheres and adding in the hard AE part of the density within the spheres. Using the above expressions for the total density partitioned into its constituent parts, it is possible to write down an expression for the AE DFT ground-state energy as a functional of $\mathbf{P}^{\{v\}}$, and it can be shown that this energy can be partitioned into augmentation sphere parts that can be treated on a radial grid around the atoms and soft PS parts that can be evaluated on the entire real space grid.

However, when partitioning the density into parts according to equation (5.26) one does have to consider energy terms due to the non-local Hartree potential of one augmentation sphere interacting with another. These terms are unwanted, as it is desirable to keep in-sphere contributions fully local. Thus, following the conventional PAW formalism outlined in references [110, 111], a compensation density $\hat{\rho}(\mathbf{r})$ is added to $\tilde{\rho}(\mathbf{r})$ and subtracted from $\tilde{\rho}^{(1)}(\mathbf{r})$. The compensation density can be calculated through

$$\hat{\rho}(\mathbf{r}) = \sum_{ijLM} \rho^{\{v\}ij} \hat{Q}_{ij}^{LM}(\mathbf{r}) \tag{5.30}$$

where $\hat{Q}_{ij}^{LM}(\mathbf{r})$ is defined in a way to ensure that

$$\int_{\Omega_R} \left[\rho^{(1)}(\mathbf{r}) - \tilde{\rho}^{(1)}(\mathbf{r}) + \hat{\rho}(\mathbf{r})\right] |\mathbf{r} - \mathbf{R}|^L S_{LM}(\mathbf{r} - \mathbf{R}) \, \mathrm{d}^3 r = 0 \tag{5.31}$$

for all L and M and the integral is defined over the augmentation region of the atom at $\mathbf{R}$. S_{LM} again denotes a spherical harmonic. This condition is fulfilled by writing $\hat{Q}_{ij}^{LM}(\mathbf{r})$

as

$$\hat{Q}_{ij}^{LM}(\mathbf{r}) = n_{n_i l_i, n_j l_j}^{L} g^{L}(\mathbf{r}) \mathcal{G}_{l_i m_i, l_j m_j}^{LM} S_{LM}(\hat{\mathbf{r}}) \tag{5.32}$$

where $\mathcal{G}_{l_i m_i, l_j m_j}^{LM}$ denotes precomputed Gaunt coefficients that are obtained by performing an integral over the product of three different spherical harmonics, $n_{n_i l_i, n_j l_j}^{L}$ is given by an integral over the radial part of the PS and AE partial waves

$$n_{n_i l_i, n_j l_j}^{L} = \int_0^{r_c} \mathrm{d}r \, \left(\varphi_{n_i l_i}(r) \varphi_{n_j l_j}(r) - \tilde{\varphi}_{n_i l_i}(r) \tilde{\varphi}_{n_j l_j}(r) \right) r^L \tag{5.33}$$

and $g^L(\mathbf{r})$ is an arbitrary shape function chosen to be a pair of Bessel functions of order L.

The above condition ensures that the Hartree potential of the difference between the soft and the AE sphere density, when augmented with the compensation density, is zero everywhere outside the augmentation region. It then becomes possible to write the interacting DFT energy (omitting the ion-ion energy) in terms of soft PS quantities that can be treated on the real space grid, and soft and AE sphere terms such that

$$E = \tilde{E} - \tilde{E}^{(1)} + E^{(1)} \tag{5.34}$$

where[25]

$$\begin{aligned}
\tilde{E} &= P^{\{v\}\alpha\beta} \langle \phi_\alpha | -\nabla^2 | \phi_\beta \rangle + E_{\mathrm{H}} \left[\tilde{\rho} + \hat{\rho} \right] + E_{\mathrm{xc}} \left[\tilde{\rho} + \hat{\rho} \right] \\
&\quad + \int V_{\mathrm{H}} \left[\tilde{\rho}_{Z_c} \right] (\mathbf{r}) \left[\tilde{\rho} + \hat{\rho} \right] (\mathbf{r}) \mathrm{d}^3 r \qquad (5.35) \\
\tilde{E}^{(1)} &= \rho^{\{v\}ij} \langle \tilde{\varphi}_i | -\nabla^2 | \tilde{\varphi}_j \rangle + E_{\mathrm{H}} \left[\tilde{\rho}^{(1)} + \hat{\rho} \right] + E_{\mathrm{xc}} \left[\tilde{\rho}^{(1)} + \hat{\rho} \right] \\
&\quad + \int_\Omega V_{\mathrm{H}} \left[\tilde{\rho}_{Z_c} \right] (\mathbf{r}) \left[\tilde{\rho}^{(1)} + \hat{\rho} \right] (\mathbf{r}) \mathrm{d}^3 r \qquad (5.36) \\
E^{(1)} &= \sum_{ij} \rho^{\{v\}ij} \langle \varphi_i | -\nabla^2 | \varphi_j \rangle + E_{\mathrm{H}} \left[\rho^{(1)} \right] + E_{\mathrm{xc}} \left[\rho^{(1)} \right] \\
&\quad + \int_\Omega V_{\mathrm{H}} \left[\rho_{Z_c} \right] (\mathbf{r}) \rho^{(1)}(\mathbf{r}) \mathrm{d}^3 r. \qquad (5.37)
\end{aligned}$$

Here, ρ_{Z_c} denotes the AE core density of the ions, while $\tilde{\rho}_{Z_c}$ is the soft PS version of the ion core density that can be expressed on the regular simulation cell grid. For the integrals, Ω denotes the volume of the augmentation spheres only. In the spirit of conventional PAW approaches, one wants to find the energy of the system as a functional of the soft density matrix $\mathbf{P}^{\{v\}}$ and thus want to write the bandstructure energy of the system at

[25]Note that the soft compensation density $\hat{\rho}$ is not always included into E_{xc} in practical calculations, since its unphysical shape can sometimes create unwanted effects in the exchange-correlation potential. In this work, all equations are written in such a way that it is assumed that $\hat{\rho}$ is included in E_{xc}.

fixed Kohn-Sham potential in terms of an effective PAW Hamiltonian $\mathbf{H}^{\phi}$ and $\mathbf{P}^{\{v\}}$, such that

$$E_{\mathrm{BS}} = 2\mathrm{Tr}\left[\mathbf{H}^{\phi}\mathbf{P}^{\{v\}}\right]. \tag{5.38}$$

In order to find this PAW-transformed DFT Hamiltonian, note that

$$\mathbf{H}^{\phi} = \frac{\mathrm{d}E_{\mathrm{BS}}}{\mathrm{d}\mathbf{P}^{\{v\}}} = \frac{\partial E_{\mathrm{BS}}}{\partial \mathbf{P}^{\{v\}}} + \sum_{ij} \frac{\partial E_{\mathrm{BS}}}{\partial \rho^{\{v\}ij}} \frac{\partial \rho^{\{v\}ij}}{\partial \mathbf{P}^{\{v\}}} + \int \mathrm{d}^3 r \, \frac{\delta E_{\mathrm{BS}}}{\delta \hat{\rho}(\mathbf{r})} \frac{\partial \hat{\rho}(\mathbf{r})}{\partial \mathbf{P}^{\{v\}}} \tag{5.39}$$

and thus the effective Hamiltonian can be constructed in a straightforward manner by differentiating the AE bandstructure energy. It can be shown [110, 111] that the PAW Hamiltonian $\mathbf{H}^{\phi}$ can be written as

$$\left(H^{\phi}\right)_{\alpha\beta} = \langle\phi_{\alpha}| - \frac{1}{2}\nabla^2 + \tilde{V}_{\mathrm{KS}}|\phi_{\beta}\rangle + \sum_{ij} Q^i_{\alpha} D_{ij} \left(Q^j_{\beta}\right)^{\dagger} \tag{5.40}$$

where the effective soft part of the Kohn–Sham potential $\tilde{V}_{\mathrm{KS}}$ is given by

$$\tilde{V}_{\mathrm{KS}}(\mathbf{r}) = V_{\mathrm{H}}\left[\tilde{\rho} + \hat{\rho} + \tilde{\rho}_{Z_c}\right](\mathbf{r}) + V_{\mathrm{xc}}\left[\tilde{\rho} + \hat{\rho} + \tilde{\rho}_c\right]. \tag{5.41}$$

Here, $\tilde{\rho}_c$ is the soft PS density of the core electrons inside the augmentation region. Thus the total PAW Hamiltonian can be written in terms of a PS Kohn–Sham Hamiltonian constructed on the real space grid, and an augmentation sphere correction term defined through D_{ij}. The PAW correction term D_{ij} contains three major contributions. The first one is a density independent term due to the kinetic energy operator acting on the AE and PS partial waves which only has to be constructed once at the beginning of the calculation. The two other terms originate from the Hartree and exchange-correlation potential and are density dependent. Thus a new D_{ij} has to be constructed every time the density defining the Kohn–Sham Hamiltonian changes.

The evaluation of the Hartree and exchange-correlation PAW contributions inside the augmentation spheres is facilitated by expanding both the Coulomb potential and the exchange-correlation potential in terms of the spherical harmonics $S_{LM}(\hat{\mathbf{r}})$. Given that the AE and PS partial waves are expanded in the same way, the sphere terms can then be expressed in terms of a summation over Gaunt coefficients $\mathcal{G}^{LM}_{l_i m_i, l_j m_j}$ and a radial integral. Using this approach, it can be shown that the full PAW correction term can be written as

$$D_{ij} = D^0_{ij} + \hat{D}_{ij} + \sum_{kl} \rho^{\{v\}kl} e_{ijkl} + D^{\mathrm{xc}}_{ij} \tag{5.42}$$

where e_{ijkl} is a rank-4 tensor that can be precomputed at the beginning of a PAW calculation from the PAW partial waves and that contains all PAW corrections originating

from the Hartree term. D^0_{ij} is the constant term due to the kinetic energy operator acting on the AE and PS partial waves. $\hat{D}_{ij}$ is closely related to the compensation density and can be easily evaluated from the soft Kohn–Sham potential $\tilde{V}_{\mathrm{KS}}$ via[26]

$$\hat{D}_{ij} = \sum_{LM} \int \mathrm{d}^3 r \, \tilde{V}_{\mathrm{KS}}(\mathbf{r}) \hat{Q}^{LM}_{ij}(\mathbf{r}) \tag{5.43}$$

The exchange-correlation term D^{xc}_{ij} is more difficult to construct as it requires the evaluation of $V^{LM}_{\mathrm{xc}}(r)$ inside the PAW augmentation spheres given by

$$V^{LM}_{\mathrm{xc}}(r) = \int \mathrm{d}\sigma \, V_{\mathrm{xc}}[\rho](\mathbf{r}) S_{LM}(\hat{\mathbf{r}}). \tag{5.44}$$

Here, $\mathrm{d}\sigma$ is taken to denote the differential solid angle. In ONETEP (following the implementation discussed in [111]), the term is evaluated by assuming that close to the atom, the density is close to spherically symmetric and thus the full density can be expressed by a Taylor expansion to second order around the symmetric density.

Thus in general, the linear-scaling DFT algorithm remains unchanged by the PAW formalism. The AE DFT energy is minimised by simply optimising the soft PS density matrix $\mathbf{P}^{\{v\}}$ obeying the usual normalisation and idempotency condition with a redefined overlap matrix $\mathbf{S}^{\phi}$. Every time the Hamiltonian is constructed, one simply has to calculate the nonlocal correction term D_{ij} and add the term $(V_{nl})_{\alpha\beta} = \sum_{ij} Q^i_\alpha D_{ij} \left(Q^j_\beta\right)^\dagger$ to the soft Hamiltonian matrix evaluated on the entire simulation cell grid. Since the correction term is evaluated inside the augmentation spheres only and the compensation density guarantees that the Hartree potential in one sphere does not contribute to another sphere, the matrix $(V_{nl})_{\alpha\beta}$ does not destroy any of the sparsity properties of the original pseudopotential Hamiltonian. Since D_{ij} can be computed in linear-scaling effort, it follows that the PAW density-matrix formalism used in ONETEP preserves the linear-scaling properties of the original method.

5.2.3 The implicit solvent model

The large system sizes accessible in linear-scaling DFT methods have allowed for biological systems to become increasingly of interest to the DFT community. In these systems, the question of solvent effects becomes very important. Many structures of proteins obtained via X-ray crystallography are obtained within some form of solvating medium. Taking these structures and simply simulating them in vacuum can yield unphysical vanishing

[26]Note that since the soft Kohn–Sham potential is defined on the real space grid, this integral is written as an integral over the entire real space grid. However, since $\hat{Q}^{LM}_{ij}(\mathbf{r})$ is only nonzero inside the augmentation regions, the integral can be evaluated in practice by considering a small augmentation box around the atom on the real space grid.

band-gaps and convergence problems caused by unscreened electrostatic charges on the surface of the protein [112]. Including the solvent environment explicitly in the calculation, either via classical point charges or fully quantum mechanically, is possible but in the case of classical charges introduces uncontrolled approximations and in the case of fully quantum mechanical solvents significantly increases the computational complexity. An alternative approach is to introduce the effects of the screening medium into the calculation via an implicit solvent model (see [113] for a review of different approaches to the implicit solvent model encountered in the quantum chemistry community), which has been recently implemented in `ONETEP` [114].

In general, the implicit solvent model is defined through a permittivity $\epsilon(\mathbf{r})$ such that

$$\epsilon\left[\rho\right](\mathbf{r}) = 1 + \frac{\epsilon_\infty - 1}{2}\left(1 + \frac{1 - (\rho(\mathbf{r})/\rho_{\text{iso}})^{2\beta}}{1 + (\rho(\mathbf{r})/\rho_{\text{iso}})^{2\beta}}\right) \tag{5.45}$$

where ϵ_∞ is the bulk permittivity and ρ_{iso} defines the cavity of the solute where the bulk permittivity drops to $\epsilon_\infty/2$. There are several different ways of defining the solute cavity, such as taking the effective surface generated by placing spheres of a fixed radius onto the atomic elements of the solute [113][27]. For the purpose of this work however, ρ_{iso} is taken to be an isosurface of the ground state density. This has the advantage that the solute cavity can in principle respond self-consistently to the solvent. Equation (5.45) thus defines a permittivity smoothly changing from the bulk permittivity value to 1 inside a cavity defined by the density of the system and β is a parameter defining the smoothness of the transition.

The Hartree terms of the electron-electron and electron-ion interaction in the Kohn–Sham DFT potential are then replaced by the total electrostatic potential $\Phi_{\text{ES}}(\mathbf{r})$, which is obtained by solving the Poisson equation

$$\nabla \cdot \left(\epsilon\left[\rho\right]\nabla\Phi_{\text{ES}}\right) = -4\pi\rho_{\text{tot}} \tag{5.46}$$

in real space subject to open boundary conditions. Here, ρ_{tot} is the total density of ρ and a Gaussian-smeared density of the ionic cores of the atoms.

Since the cavity is defined as an isosurface of the density, the gradient of the total electrostatic energy with respect to the Kohn–Sham density is no longer simply given by

[27] As is pointed out in [113], the freedom of choice in defining the solute cavity is of little consequence in practical calculations, as the implicit solvent model is, by definition, an approximate approach to account for solvation effects. In situations where the immediate surroundings of the solute need to be treated more precisely, it is always possible to introduce a first solvantion shell of explicit solvent molecules, followed by an implicit solvent model to treat the continuum.

$\Phi_{\mathrm{ES}}(\mathbf{r})$ but contains a term due to changes in the isosurface:

$$\frac{\delta E_{\mathrm{ES}}}{\delta \rho}(\mathbf{r}) = \Phi_{\mathrm{ES}}(\mathbf{r}) - \frac{1}{8\pi}\left(\nabla \Phi_{\mathrm{ES}}(\mathbf{r})\right)^2 \frac{\delta \epsilon}{\delta \rho}(\mathbf{r}) \tag{5.47}$$

The second term is difficult to treat numerically, as $\frac{\delta \epsilon}{\delta \rho}(\mathbf{r}) \to 0$ everywhere apart from close to the boundary defined by the density isosurface, which is precisely the region where $\left(\nabla \Phi_{\mathrm{ES}}(\mathbf{r})\right)^2$ becomes very small. In practice, this self-consistent response of the solvation cavity is often ignored to remove numerical ill-conditioning. This is normally achieved by performing a ground state DFT calculation in vacuum and open boundary conditions and then fixing the cavity in the in-solvent calculation to an isosurface of the ground-state density obtained in vacuum. In general, the fixed cavity approximation is of little consequence to the calculated free energies of solvation of neutral molecules [114], and is thus suitable for the systems of interest considered in this work.

5.2.4 The electronic enthalpy method

In many situations it is desirable to study a system under transformation due to an externally applied pressure. In a DFT framework, this can be straightforwardly achieved by considering not the Kohn–Sham ground-state energy, but the enthalpy as a property to be minimised, such that

$$H\left[\mathbf{P}^{\{v\}}, \{\phi_\alpha\}, \{\mathbf{R}_\gamma\}\right] = E_{\mathrm{KS}}\left[\mathbf{P}^{\{v\}}, \{\phi_\alpha\}, \{\mathbf{R}_\gamma\}\right] + PV\left[\{\mathbf{R}_\gamma\}\right] \tag{5.48}$$

where P and V are the pressure and volume of the system and $\{\mathbf{R}_\gamma\}$ are the atomic positions. For a periodic system, V is a well defined quantity and can be taken to be the volume of the unit cell. The above enthalpy can then be minimised at constant pressure subject to the atomic positions, which yields the structure of the system under pressure P. Calculating the crystal structure of the system under a range of different pressures yields information on phase transitions in the studied material.

While the above method is well defined for a periodic crystal, the thermodynamic enthalpy becomes an ill-defined quantity for finite or zero-dimensional systems like quantum dots, as there is no longer a unique way of defining the volume V. One way to study such systems is to include the pressurising medium explicitly in the calculation, which is computationally expensive. A conceptually simple alternative approach was introduced by Cococcioni [115] and coworkers and adapted to the `ONETEP` code by Corsini *et al* [116]. In their approach, the volume of the finite system is defined via an isosurface of the

Kohn–Sham ground state density such that

$$V = \int \mathrm{d}^3 r \, \frac{1}{2} \mathrm{erfc} \left(\frac{\alpha - \rho(\mathbf{r})}{\sigma\sqrt{2}} \right) \tag{5.49}$$

where α is the chosen cutoff defining the density isosurface and σ is a smearing parameter used for numerical reasons. Thus the PV-term in the enthalpy method becomes a functional of the electronic ground-state density only and its effect is transmitted to the ions via forces obtained through the Hellman-Feynman theorem [116]. Finding the gradient of the PV term with respect to the density leads to an effective potential contribution

$$\Phi_{PV}(\mathbf{r}) = \frac{P}{\sigma\sqrt{2\pi}} \exp \left(-\frac{(\rho(\mathbf{r}) - \alpha)^2}{2\sigma^2} \right) \tag{5.50}$$

that can be simply added to the usual Kohn–Sham potential in the DFT energy minimisation, leaving the general algorithm unchanged.

There is clearly some freedom in the choice of parameters α and σ, resulting in the fact that the effective pressure felt by the system is not necessarily equal to P. Generally α has to be chosen in a sensible way so that the isosurface of the density provides a relatively smooth surface without any cavities. However, very small values of α lead to a large excluded region and potential tails of Φ_V that fail to vanish far away from the system. Generally, the choice of α and σ needs to be calibrated for every system that is studied. This can be achieved by comparing the nanocrystal bulk modulus for different choices of P with the bulk modulus of an infinite system at the same pressures.

Chapter 6

Linear-scaling TDDFT in ONETEP

In this section, an approach to solve for the lowest eigenstates of the TDDFT linear-response equation in linear scaling effort is outlined and its implementation in the ONETEP code is discussed in some detail. For the purpose of this discussion the treatment is limited to the Tamm–Dancoff approximation and closed shell systems. $\mathbf{H}_{\mathrm{TDA}}^{[2p]}$ again denotes the effective Hermitian two-particle Hamiltonian in canonical Kohn–Sham state representation (see section 4.2.5) and $\mathbf{x}$ is taken to be a trial eigenstate of $\mathbf{H}_{\mathrm{TDA}}^{[2p]}$ in the product space $\{|\psi_v\rangle \otimes |\psi_c\rangle\}$. Then the lowest eigenstate ω of the system can be written in terms of a variational principle by minimising the Rayleigh–Ritz eigenvalue with respect to the trial vector $\mathbf{x}$:

$$\omega = \min_{\mathbf{x}} \frac{\mathbf{x}^\dagger \mathbf{H}_{\mathrm{TDA}}^{[2p]} \mathbf{x}}{\mathbf{x}^\dagger \mathbf{x}}. \tag{6.1}$$

Differentiating (6.1) with respect to $\mathbf{x}$, one obtains

$$\frac{\partial \omega}{\partial \mathbf{x}} = \frac{2}{\mathbf{x}^\dagger \mathbf{x}} \left[\mathbf{H}_{\mathrm{TDA}}^{[2p]} \mathbf{x} - \omega \mathbf{x} \right]. \tag{6.2}$$

The gradient of the Rayleigh-Ritz value with respect to changes in $\mathbf{x}$ can then be used in conjugate gradient algorithms to iteratively solve for the lowest eigenstate of the system. Note that for this iterative solution, only knowledge of the action $\mathbf{q} = \mathbf{H}_{\mathrm{TDA}}^{[2p]} \mathbf{x}$ is required, and no explicit calculation of the entire matrix $\mathbf{H}_{\mathrm{TDA}}^{[2p]}$ is required, which can only be obtained in $\mathcal{O}(N^4)$ computational effort. In Kohn–Sham product space, the vector $\mathbf{q}$ describing this action can be expressed as[28]

$$\begin{aligned} q_{cv} &= \sum_{c'v'} \left(H_{\mathrm{TDA}}^{[2p]} \right)_{cv,c'v'} x_{c'v'} = (\epsilon_c^{\mathrm{KS}} - \epsilon_v^{\mathrm{KS}}) x_{cv} \\ &\quad + 2 \sum_{c'v'} \left\langle \psi_c^* \psi_v \left| \frac{1}{|\mathbf{r} - \mathbf{r}'|} + \left. \frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}')} \right|_{\rho_0} \right| \psi_{c'} \psi_{v'}^* \right\rangle x_{c'v'}. \end{aligned} \tag{6.3}$$

[28] for the purpose of this work, any memory effects in the exchange-correlation kernel is ignored and thus $\mathbf{H}_{\mathrm{TDA}}^{[2p]}$ is taken to be frequency independent.

As it can be seen, the action splits into the diagonal term of Kohn–Sham eigenvalue differences and a coupling term resulting from the Hartree and exchange-correlation interaction.

Following Hutter [117], it is now helpful to introduce the first-order response density $\rho^{\{1\}}$ that is associated with trial vector $\mathbf{x}$ such that[29] :

$$\rho^{\{1\}}(\mathbf{r}) = \sum_{c,v} \langle \mathbf{r}|\psi_c\rangle x_{cv} \langle \psi_v|\mathbf{r}\rangle \tag{6.4}$$

Defining the self-consistent field potential $V^{\{1\}}_{\mathrm{SCF}}[\rho^{\{1\}}](\mathbf{r})$ as a reaction to the response density as

$$\begin{aligned} V^{\{1\}}_{\mathrm{SCF}}[\rho^{\{1\}}](\mathbf{r}) &= 2\int \mathrm{d}^3r' \, \frac{\rho^{\{1\}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &+ 2\int \mathrm{d}^3r' \, \frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\bigg|_{\rho_0} \rho^{\{1\}}(\mathbf{r}') \end{aligned} \tag{6.5}$$

the action $\mathbf{q}$ of the TDDFT operator $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$ on the arbitrary trial vector $\mathbf{x}$ can be simply written as

$$q_{cv} = (\epsilon^{\mathrm{KS}}_c - \epsilon^{\mathrm{KS}}_v)x_{cv} + \langle \psi_c|V^{\{1\}}_{\mathrm{SCF}}[\rho^{\{1\}}](\mathbf{r})|\psi_v\rangle. \tag{6.6}$$

If $\mathbf{q}$ can be constructed from Equation (6.6) in linear-scaling effort for arbitrary $\mathbf{x}$ then the lowest excitation of the system can likewise be found in $\mathcal{O}(N)$ operations. It should be pointed out that this approach of finding the lowest eigenstate is based on the variational principle and thus is only valid if the eigenproblem is forced to be Hermitian due to the Tamm–Dancoff approximation. However, the full TDDFT Hamiltonian consists of Hermitian blocks and thanks to this structure it is possible to form a more generalised variational principle valid for the full problem [118]. Generally speaking, the full eigenvalue problem is more difficult to solve than the Hermitian equivalent under the Tamm–Dancoff approximation and is considered to be outside the scope of this work.

6.1 Derivation of the formalism

In canonical representation, $\mathbf{q}$ makes explicit reference to eigenergies of occupied and unoccupied Kohn–Sham bands, which cannot be obtained in linear-scaling effort. In order to obtain a formalism in which $\mathbf{q}$ can be constructed in $\mathcal{O}(N)$ operations, any reference to individual Kohn–Sham eigenstates must be avoided. Reference can only be made to

[29] Note that $\rho^{\{1\}}$ is not required to be real and will generally not be real, if $\{\psi_i\}$ cannot be chosen to be real. On the other hand, $\sum_{cv} x^\dagger_{cv}\langle\psi_c|V^{\{1\}}_{\mathrm{SCF}}[\rho^{\{1\}}](\mathbf{r})|\psi_v\rangle$ is necessarily real and thus ω is real as required. In the `ONETEP` method however, calculations are performed at the τ-point and thus $\rho^{\{1\}}$ is trivially real.

collections of Kohn–Sham states in the form of single-particle density matrices defined over some subspace of Kohn–Sham bands. Furthermore, local orbitals have to be used in order to represent $\rho^{\{1\}}$ and $V_{\mathrm{SCF}}^{\{1\}}[\rho^{\{1\}}](\mathbf{r})$ to exploit locality when evaluating matrix elements of $\mathbf{q}$. The two factors of localised representations of all involved quantities and the use of a Kohn–Sham density-matrix formalism throughout can be seen as the main ingredients to linear-scaling TDDFT [98–100]. In the following outline of the formalism, the original approach in deriving the method for the ONETEP code published in [99] is closely followed.

6.1.1 The TDDFT operator in $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ representation

A ground state and conduction state calculation in ONETEP as described in chapter 5 yields two sets of localised orbitals, $\{\phi_\alpha\}$ and $\{\chi_\beta\}$, as well as two single-particle Kohn–Sham density matrices $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{c\}}$ spanning the valence and low energy conduction subspace respectively. The localised orbitals are optimised to form an ideal representation of their respective subspaces such that only a minimal number is required, making the representation very small and efficient. Therefore in combination, $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ form a suitable representation to expand quantities like $\rho^{\{1\}}$ and $V_{\mathrm{SCF}}^{\{1\}}[\rho^{\{1\}}]$.

Starting with the response density, and again using superscript and subscript greek indices to differentiate between contravariant and covariant tensor quantities respectively, one can write

$$\rho^{\{1\}}(\mathbf{r}) = \sum_{c,v} \langle \mathbf{r}|\psi_c^{\mathrm{KS}}\rangle x_{cv} \langle \psi_v^{\mathrm{KS}}|\mathbf{r}\rangle = \sum_{v}^{\mathrm{occ}} \sum_{c}^{\mathrm{opt}} \langle \mathbf{r}|\chi_\alpha\rangle \langle \chi^\alpha|\psi_c^{\mathrm{KS}}\rangle x_{cv} \langle \psi_v^{\mathrm{KS}}|\phi^\beta\rangle \langle \phi_\beta|\mathbf{r}\rangle. \tag{6.7}$$

Here, the sum of the conduction states goes over all the states for which $\{\chi_\alpha\}$ is optimised. The approximation of limiting the sum to states for which $\{\chi_\alpha\}$ is explicitly optimised goes beyond the implicit approximation involved when $\rho^{\{1\}}$ is expanded in a finite-size representation. Any finite-size representation expansion automatically limits the infinite unoccupied subspace to a finite number of N_{basis} Kohn–Sham bands, where N_{basis} is the number of basis functions in the finite representation. By limiting the sum to only optimised states, it is ensured that the only states included in the calculation are well-represented by the chosen double representation of $\{\chi_\alpha\}$ and $\{\phi_\beta\}$, a condition that is generally not obeyed by other finite size representations.

In order for the approximation to be valid it is required that $\rho^{\{1\}}$ is well described by a relatively small number of unoccupied states for some low energy excitations of interest in the chosen system. This approximation can be rigorously tested by including a larger subset of the conduction space manifold in the optimisation of the conduction density matrix $\mathbf{P}^{\{c\}}$. In the spirit of the linear-scaling DFT formalism one can now introduce the

response density matrix $\rho^{\{1\}}(\mathbf{r},\mathbf{r}')$ such that

$$\rho^{\{1\}}(\mathbf{r},\mathbf{r}) = \rho^{\{1\}}(\mathbf{r}) = \chi_\alpha(\mathbf{r}) P^{\{1\}\alpha\beta} \phi_\beta(\mathbf{r}) \tag{6.8}$$

where the effective response density matrix $P^{\{1\}\alpha\beta}$ in mixed NGWF representation is defined as

$$P^{\{1\}\alpha\beta} = \sum_{v}^{\text{occ}} \sum_{c}^{\text{opt}} \langle \chi^\alpha | \psi_c^{\text{KS}} \rangle x_{cv} \langle \psi_v^{\text{KS}} | \phi^\beta \rangle. \tag{6.9}$$

Thus provided that $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ are capable of representing $\rho^{\{1\}}(\mathbf{r},\mathbf{r}')$ for a given low energy excitation, $P^{\{1\}\alpha\beta}$ is a representation of the TDDFT trial vector $\mathbf{x}$ in the product vector space $\{|\chi_\alpha\rangle \otimes |\phi_\beta\rangle\}$ spanned by the non-orthogonal localised orbitals rather than the product vector space $\{|\psi_v\rangle \otimes |\psi_c\rangle\}$ of orthonormal Kohn–Sham states.

Similarly to the response density, the coupling matrix elements of $\mathbf{q}$, $(V_{\text{SCF}}^{\{1\}})_{cv} = \langle \psi_c | V_{\text{SCF}}^{\{1\}}[\rho^{\{1\}}] | \psi_v \rangle$ can be rewritten as

$$(V_{\text{SCF}}^{\{1\}})_{cv} = \langle \psi_c^{\text{KS}} | \chi^\alpha \rangle \langle \chi_\alpha | \hat{V}_{\text{SCF}}^{\{1\}} | \phi_\beta \rangle \langle \phi^\beta | \psi_v^{\text{KS}} \rangle. \tag{6.10}$$

Furthermore, the diagonal part of q_{cv} consisting of Kohn–Sham conduction-valence eigenvalue differences becomes:

$$\begin{aligned}(\epsilon_c^{\text{KS}} - \epsilon_v^{\text{KS}}) x_{cv} = &\sum_{c'}^{\text{opt}} \langle \psi_c^{\text{KS}} | \chi^\alpha \rangle \langle \chi_\alpha | \hat{H} | \chi_\beta \rangle \langle \chi^\beta | \psi_{c'}^{\text{KS}} \rangle x_{c'v} \\ &- \sum_{v'}^{\text{occ}} x_{cv'} \langle \psi_{v'}^{\text{KS}} | \phi^\alpha \rangle \langle \phi_\alpha | \hat{H} | \phi_\beta \rangle \langle \phi^\beta | \psi_v^{\text{KS}} \rangle.\end{aligned} \tag{6.11}$$

Note that the matrix elements involving the Kohn–Sham Hamiltonian operator simply correspond to the Hamiltonian matrices $\mathbf{H}^\chi$ and $\mathbf{H}^\phi$ in their respective conduction and valence NGWF representation. The self consistent field response in mixed conduction-valence NGWF representation is denoted as $\mathbf{V}_{\text{SCF}}^{\{1\}\chi\phi}[\rho^{\{1\}}]$ such that

$$(V_{\text{SCF}}^{\{1\}\chi\phi}[\rho^{\{1\}}])_{\alpha\beta} = \langle \chi_\alpha | V_{\text{SCF}}^{\{1\}}[\rho^{\{1\}}] | \phi_\beta \rangle. \tag{6.12}$$

By inserting equations (6.11) and (6.10) into equation (6.6), multiplying with $\langle \chi^\alpha | \psi_c^{\text{KS}} \rangle$ and $\langle \psi_v^{\text{KS}} | \phi^\beta \rangle$ from the left and right respectively and summing over the c and v indices, one can remove all references to the canonical representation from $\mathbf{q}$. Using the definition of the response density matrix $\mathbf{P}^{\{1\}}$, the result of the TDDFT operator acting on a trial

response matrix in NGWF representation reduces to the simple form

$$\begin{aligned}(q^{\chi\phi})^{\alpha\beta} &= (P^{\{\mathrm{c}\}}H^{\chi}P^{\{1\}} - P^{\{1\}}H^{\phi}P^{\{\mathrm{v}\}})^{\alpha\beta} \\ &+ (P^{\{\mathrm{c}\}}V_{\mathrm{SCF}}^{\{1\}\chi\phi}[\rho^{\{1\}}]P^{\{\mathrm{v}\}})^{\alpha\beta}. \end{aligned} \tag{6.13}$$

Note that (6.13) makes no reference to individual Kohn–Sham states or bandstructure energies. The only reference that is made is to collective representations in the form of the density matrices of the valence and low energy conduction space in NGWF representation, as well as the Kohn–Sham Hamiltonian in the respective NGWF representations. $\mathbf{H}^{\chi}$, $\mathbf{H}^{\phi}$, $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{v\}}$ can all be obtained in linear scaling effort by performing a ground state and conduction state optimisation in ONETEP. Evaluating the first two terms in (6.13) for a given $\mathbf{P}^{\{1\}}$ then reduces to a number of matrix multiplications, where the dimensions of the matrices grow as $\mathcal{O}(N)$ with system size. Due to the strict localisation of the NGWF representations, constructing $\mathbf{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}[\rho^{\{1\}}]$ for a given response potential $V_{\mathrm{SCF}}^{\{1\}}[\rho^{\{1\}}](\mathbf{r})$ requires $\mathcal{O}(N)$ computational steps and the matrix $\mathbf{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}[\rho^{\{1\}}]$ necessarily becomes sparse in the limit of large systems. The scaling of the calculation of the response potential $V_{\mathrm{SCF}}^{\{1\}}[\rho^{\{1\}}](\mathbf{r})$ depends on the nature of the exchange-correlation functional used, but is necessarily linear for all semi-local exchange-correlation functionals.

6.1.2 The Rayleigh–Ritz value

Using the action $\mathbf{q}^{\chi\phi}$ of $\mathbf{P}^{\{1\}}$ acting on the full two-particle TDDFT Hamiltonian $\mathbf{H}^{[2p]}$, the lowest excitation energy can be obtained from optimising

$$\omega = \min_{\mathbf{P}^{\{1\}}} \left\{ \frac{\mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{q}^{\chi\phi}\mathbf{S}^{\phi}\right]}{\mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\right]} \right\}. \tag{6.14}$$

The above expression is trivially shown to be equivalent to optimising the Rayleigh–Ritz value in canonical representation. However, the minimistation in (6.14) is not valid for any arbitrary variation of the response density matrix and thus unlike in the canonical Kohn–Sham representation, there is an implied constraint on $\mathbf{P}^{\{1\}}$. This constraint originates from the way $\mathbf{P}^{\{1\}}$ is defined in (6.9) and can be expressed in form of an invariance relationship:

$$\mathbf{P}^{\{1\}'} = \mathbf{P}^{\{\mathrm{c}\}}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{\mathrm{v}\}} = \mathbf{P}^{\{1\}}. \tag{6.15}$$

The constraint is analogous to the idempotency constraint on $\mathbf{P}^{\{v\}}$ in density matrix based DFT. Indeed, it originates from expanding the idempotency constraint on perturbations to the ground state density matrix to first order [119]. Density matrices that violate (6.15) do not describe excitations from occupied to unoccupied states, but also mix in

forbidden occupied-occupied and virtual-virtual transitions. Just like in ground state DFT, it becomes necessary to enforce the idempotency constraint explicitly as soon as the canonical Kohn–Sham state representation is abandonned for a more flexible one. In contrast, in the minimisation in (6.1), no constraint is placed on variations of $\mathbf{x}$ as $\{|\psi_c\rangle\}$ and $\{|\psi_v\rangle\}$ are orthonormal to each other.

In practice, enforcing the condition in (6.15) during the minimisation of the excitation energy does not cause the same problems as in ground state DFT. To see why, consider minimising (6.14) by finding the derivative of the Rayleigh-Ritz eigenvalue with respect to $\mathbf{P}^{\{1\}}$:

$$\frac{\partial\omega}{\partial\mathbf{P}^{\{1\}}} = \mathbf{g}^{\chi\phi} = \frac{2}{\mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\right]}\left[\mathbf{S}^{\chi}\mathbf{q}^{\chi\phi}\mathbf{S}^{\phi} - \omega\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\right]. \tag{6.16}$$

$\mathbf{g}^{\chi\phi}$ is the covariant gradient of the TDDFT excitation energy with respect to the contravariant density matrix $\mathbf{P}^{\{1\}}$. To transform it into a contravariant quantity, one multiplies from the left and the right with the appropriate inverse overlap matrices such that

$$\mathbf{f}^{\chi\phi} = \left(\mathbf{S}^{\chi}\right)^{-1}\mathbf{g}^{\chi\phi}\left(\mathbf{S}^{\phi}\right)^{-1} = \frac{2}{\mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\right]}\left[\mathbf{q}^{\chi\phi} - \omega\mathbf{P}^{\{1\}}\right]. \tag{6.17}$$

$\mathbf{f}^{\chi\phi}$ can be used as a steepest descent search direction for a conjugate gradient algorithm and has the same tensorial properties as $\mathbf{P}^{\{1\}}$. Note however, that for a response density matrix obeying the invariance constraint (6.15), $\mathbf{q}^{\chi\phi}$ obeys the same constraint by construction, as equation (6.13) is invariant under that transformation as long as $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{c\}}$ are idempotent. From this it follows that $\mathbf{f}^{\chi\phi}$ is also invariant under the operation (6.15) and so is any conjugate gradient search direction derived from it. Thus optimising a matrix $\mathbf{P}^{\{1\}}$ invariant under the transformation along a search direction derived from $\mathbf{f}^{\chi\phi}$ will yield a new $\mathbf{P}^{\{1\}}$ obeying the constraint. When minimising (6.14), it is therefore sufficient to have an initial density matrix $\mathbf{P}^{\{1\}}_{\mathrm{init}}$ that obeys the invariance relation. This invariant $\mathbf{P}^{\{1\}}_{\mathrm{init}}$ however is easily constructed from a random starting guess matrix $\mathbf{P}^{\{1\}}_{\mathrm{guess}}$ via a single projection into the occupied and unoccupied subspace:

$$\mathbf{P}^{\{1\}}_{\mathrm{init}} = \mathbf{P}^{\{\mathrm{c}\}}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}_{\mathrm{guess}}\mathbf{S}^{\phi}\mathbf{P}^{\{\mathrm{v}\}}. \tag{6.18}$$

6.1.3 Post-processing analysis

Once (6.14) is minimised, $\mathbf{P}^{\{1\}}$ is the response density matrix corresponding the lowest excitation of the system with associated excitation energy ω. From this response density matrix, a number of interesting physical properties can be evaluated. The main quantity of interest is the oscillator strength that determines how strongly the transition couples

to light. It can be straightforwardly evaluated in linear-scaling effort by making use of

$$f_\omega = \frac{2\omega}{3} \left| P^{\{1\}\alpha\beta} \langle \phi_\beta | \mathbf{r} | \chi_\alpha \rangle \right|^2 . \tag{6.19}$$

Furthermore, $\mathbf{P}^{\{1\}}$ can be used to construct both electron and hole densities of the excitation via

$$\rho^{\{\text{elec}\}}(\mathbf{r}) = \chi_\alpha(\mathbf{r}) \left(P^{\{1\}} S^\phi P^{\{1\}\dagger} \right)^{\alpha\beta} \chi_\beta(\mathbf{r}) \tag{6.20}$$

$$\rho^{\{\text{hole}\}}(\mathbf{r}) = \phi_\alpha(\mathbf{r}) \left(P^{\{1\}\dagger} S^\chi P^{\{1\}} \right)^{\alpha\beta} \phi_\beta(\mathbf{r}). \tag{6.21}$$

While representing $\mathbf{P}^{\{1\}}$ in the space spanned by the conduction and valence NGWFs allows for very efficient calculations compared to the canonical representation, the transition density matrix can now no longer be straightforwardly interpreted in terms of Kohn–Sham transitions. However, decomposing TDDFT transitions in terms of Kohn–Sham transitions often forms an important diagnostic tool in TDDFT calculations. While this analysis cannot be performed in linear scaling effort, it can be carried out in an $\mathcal{O}(N^3)$ computational effort as a post-processing step after $\mathbf{P}^{\{1\}}$ is converged. Let $\mathbf{U}^{\{c\}}$ and $\mathbf{U}^{\{v\}}$ denote the matrices of conduction- and valence Kohn–Sham eigenvectors in NGWF space such that

$$\mathbf{U}^{\{c\}\dagger} \mathbf{H}^\chi \mathbf{U}^{\{c\}} = \mathbf{E}_{\text{KS}}^{\{c\}} \tag{6.22}$$

$$\mathbf{U}^{\{v\}\dagger} \mathbf{H}^\phi \mathbf{U}^{\{v\}} = \mathbf{E}_{\text{KS}}^{\{v\}} \tag{6.23}$$

where $\mathbf{E}_{\text{KS}}^{\{c\}}$ and $\mathbf{E}_{\text{KS}}^{\{v\}}$ are the diagonal matrices of Kohn–Sham eigenvalues. $\mathbf{U}^{\{c\}}$ and $\mathbf{U}^{\{v\}}$ can be obtained in $\mathcal{O}(N^3)$ computational effort by directly diagonalising the conduction and valence Hamiltonian. The matrices of eigenvectors can then be used to rotate $\mathbf{P}^{\{1\}}$ back into canonical Kohn–Sham space such that

$$\mathbf{x} = \mathbf{U}^{\{c\}\dagger} \mathbf{P}^{\{1\}} \mathbf{U}^{\{v\}} \tag{6.24}$$

where $\mathbf{x}$ denotes again the transition vector in Kohn–Sham space corresponding to $\mathbf{P}^{\{1\}}$. In this representation, $|x_{cv}|^2$ can then be rigorously interpreted as the percentage of the Kohn–Sham transition from Kohn–Sham state v to Kohn–Sham state c that is mixed into $\mathbf{P}^{\{1\}}$.

6.1.4 Multiple excited states

In general, one is interested in not only the single lowest, but rather the N_ω lowest excitations of a system. In that case, the outlined approach can be used to optimise a set

of response density matrices $\left\{\mathbf{P}_i^{\{1\}}; i = 1, ...N_\omega\right\}$ with corresponding $\left\{\mathbf{q}_i^{\chi\phi}; i = 1, ...N_\omega\right\}$ such that the response density matrices span the subspace of the N_ω lowest eigenstates. In order to do so one can define the function

$$\Omega = \sum_i^{N_\omega} \omega_i = \sum_i^{N_\omega} \left[\frac{\mathrm{Tr}\left[\mathbf{P}_i^{\{1\}\dagger}\mathbf{S}^\chi \mathbf{q}_i^{\chi\phi}\mathbf{S}^\phi\right]}{\mathrm{Tr}\left[\mathbf{P}_i^{\{1\}\dagger}\mathbf{S}^\chi \mathbf{P}_i^{\{1\}}\mathbf{S}^\phi\right]} \right] \tag{6.25}$$

which can be minimised with respect to $\left\{\mathbf{P}_i^{\{1\}}\right\}$ under the constraint

$$\mathrm{Tr}\left[\mathbf{P}_i^{\{1\}\dagger}\mathbf{S}^\chi \mathbf{P}_j^{\{1\}}\mathbf{S}^\phi\right] = \delta_{ij}. \tag{6.26}$$

Differentiating Ω with respect to $\mathbf{P}_i^{\{1\}}$ and assuming that $\left\{\mathbf{P}_i^{\{1\}}; i = 1, ...N_\omega\right\}$ is normalised, the steepest descent search direction $\mathbf{f}_i^{\chi\phi}$ orthogonal to all current response density matrices can be written as

$$\mathbf{f}_i^{\chi\phi} = 2\left[\mathbf{q}_i^{\chi\phi} - \sum_j \mathrm{Tr}\left[\mathbf{P}_j^{\{1\}\dagger}\mathbf{S}^\chi \mathbf{q}_i^{\chi\phi}\mathbf{S}^\phi\right]\mathbf{P}_j^{\{1\}}\right] \tag{6.27}$$

The orthogonality constraint on the response density matrices can be enforced by performing a re-orthonormalisation step at the end of each conjugate gradient iteration. For this re-orthonormalisation step, the Gram-Schmidt algorithm is used. A more detailed outline of the type of conjugate gradient algorithm used here to optimise the subspace of the lowest N_ω TDDFT eigenstates can be found in [120].

Once Ω is minimised, the individual TDDFT eigenstates and eigenenergies are obtained through a subspace diagonalisation to solve the N_ω dimensional symmetric eigenproblem

$$\mathbf{A}\mathbf{u} = \omega\mathbf{u} \tag{6.28}$$

where

$$A_{ij} = \mathrm{Tr}\left[\mathbf{P}_i^{\{1\}\dagger}\mathbf{S}^\chi \mathbf{q}_j^{\chi\phi}\mathbf{S}^\phi\right] \tag{6.29}$$

The matrix of eigenvectors $\mathbf{u}$ can then be used to build the individual TDDFT eigenvectors from the $\left\{\mathbf{P}_i^{\{1\}}; i = 1, ...N_\omega\right\}$ spanning the subspace. Due to the subspace diagonalisation, the asymptotic scaling of the method with the number of excitation energies converged is $\mathcal{O}(N_\omega^3)$. However, this computational step has a small prefactor associated with it and is only relevant for extremely large numbers of N_ω. The dominant scaling with number of excitation energies comes in the form of the Gram-Schmidt orthonormalisation procedures that have to be performed once per conjugate gradient step and that have an associated

computational complexity of $\mathcal{O}(N_\omega^2)$.

6.2 Sparsity and linear-scaling

While the above formalism makes use of localised representations of the occupied and unoccupied subspace and avoids any reference to individual Kohn–Sham states, the method does not initially scale linearly with system size, since evaluating $\mathbf{q}^{\chi\phi}$ requires a number of matrix-matrix multiplications involving $\mathbf{P}^{\{c\}}$, $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{1\}}$. As the dimensions of the density matrices involved grow as $\mathcal{O}(N)$ with system size, evaluating $\mathbf{q}^{\chi\phi}$ can only scale linearly with system size if $\mathbf{P}^{\{c\}}$, $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{1\}}$ can all be treated as sparse for sufficiently large system sizes. While the valence density matrix $\mathbf{P}^{\{v\}}$ is guaranteed to be exponentially localised in any system with a bandgap and time-reversal symmetry (see section 3.3.6), the same exponential localisation properties cannot be universally justified for $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{1\}}$ and require some more careful consideration.

The proof of exponential localisation of $\mathbf{P}^{\{v\}}$ can be extended to $\mathbf{P}^{\{c\}}$ if and only if there is a second bandgap in the conduction space manifold and $\mathbf{P}^{\{c\}}$ is taken to span the set of composite Kohn–Sham bands between the two bandgaps. Strictly speaking, this is also the only scenario in which exponentially localised Wannier functions exist for the part of the conduction manifold defined by $\mathbf{P}^{\{c\}}$ and a representation of $\mathbf{P}^{\{c\}}$ in the form of localised support functions $\{\chi_\alpha\}$ is formally justified in infinite systems (see section 5.2.1) [30]. If $\mathbf{P}^{\{c\}}$ is exponentially localised for some part of the conduction space manifold, it follows that the joint density matrix spanning both the valence bands and the composite set of bands described by $\mathbf{P}^{\{c\}}$ must also be exponentially localised. The joint density matrix can be treated as a block diagonal matrix with the valence and conduction density matrix making up the diagonal blocks. The perturbation represented by the response density matrix $\mathbf{P}^{\{1\}}$ corresponds to the off-diagonal blocks of the joint density matrix. However, the application of a small perturbation cannot break the disentanglement of the joint manifold of $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{c\}}$ from the rest of the conduction manifold and thus cannot break the exponential localisation of the joint block density matrix [99] and it follows that $\mathbf{P}^{\{1\}}$ must also be exponentially localised.

The exponential localisation of an arbitrary excitation described by $\mathbf{P}^{\{1\}}$ then hinges on two conditions: that the excitation only contains Kohn–Sham contributions into the set of unoccupied states described by $\mathbf{P}^{\{c\}}$ and that $\mathbf{P}^{\{c\}}$ itself is exponentially localised. In general there is no formal reason for the existence of a set of composite Kohn–Sham states in the conduction space manifold that does not have any degeneracies in $\mathbf{k}$-space with any other unoccupied bands in arbitrary periodic systems. However, for near one-

[30] Although Wannier-like functions with strict localisation properties can in many cases be obtained for some set of bands even if there is some degeneracy with other bands not part of the set [121].

dimensional systems and molecular crystals, where low-lying bands show little dispersion, it should always be possible to choose such a composite set for some low energy subset of the unoccupied manifold. Clearly, for any excitations into high energy Kohn–Sham states, this exponentially localised description is no longer possible, as the high energy Kohn–Sham states will start to show dispersion behaviour approaching that of free electrons with $\epsilon^{\mathrm{KS}}(k) \propto |\mathbf{k}|^2$ and a disentangling of the conduction space manifold is no longer possible. This puts a limitation on the maximum energy of excitations that can be computed in linear-scaling effort with the method described here.

However, it should be pointed out that the limitations described are not unique to the method discussed here, and not unique to linear-scaling linear-response TDDFT either. Any density-matrix based linear-response TDDFT approach making use of localised orbitals will suffer from the same limitation, as the representation of the conduction space manifold with localised orbitals is only strictly justified for a set of composite bands. In general, any localised orbital method will struggle to describe high energy excitations as the electron is excited into a state that closely resembles that of a free electron. Furthermore, linear-scaling time-evolution TDDFT has to deal with similar problems, as in those methods, the time dependent response density matrix of the evolving Kohn–Sham system is treated as exponentially localised without any formal justification [97, 122].

In general, it is expected that there is a large class of systems where the truncation of all involved density matrices and thus truly linear-scaling calculations are indeed possible. There are good reasons to assume that for a large number of excitations, $\mathbf{P}^{\{1\}}$ should be considerably more sparse than the ground state and conduction space density matrices and that premise will be revisited in chapter 8.

However, for fully linear-scaling TDDFT calculations where all density matrices are sparse, there are additional complications to the algorithm described in the previous section. The main one originates from the fact that for truncated density matrices, the invariance relationship of equation (6.15) only holds approximately. The reason for this is that $\mathbf{P}^{\{1\}'}$ is required to have the same predefined sparsity pattern as $\mathbf{P}^{\{1\}}$, necessarily forcing (6.15) to be an approximate relationship. This then causes $\mathbf{q}^{\chi\phi}$ to violate the invariance relationship and the accumulation of errors can cause the minimisation scheme outlined in the previous section to fail. In the spirit of density matrix DFT approaches, one can introduce the effective positive-semidefinite penalty functional $Q[\mathbf{P}^{\{1\}}]$ such that

$$Q\left[\mathbf{P}^{\{1\}}\right] = \mathrm{Tr}\left[\left(\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi} - \mathbf{P}^{\{1\}'\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}'}\mathbf{S}^{\phi}\right)^2\right]. \tag{6.30}$$

where $\mathbf{P}^{\{1\}'}$ is defined through (6.15). If none of the density matrices are truncated and $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{v\}}$ are idempotent, $Q[\mathbf{P}^{\{1\}}]$ strictly vanishes. For truncated matrices it

can be taken as a measurement of how much the invariance constraint is violated by $\mathbf{P}^{\{1\}}$. The penalty functional can be forced to stay below a chosen threshold value, by iteratively applying the projection of (6.15) to the response density matrix at the end of each conjugate gradient iteration, thus stabilising the algorithm. It should be pointed out that the violation of the invariance relationship is, unlike the violation of the idempotency constraint in ground state density matrix DFT, an issue that arises purely from numerical errors in sparse matrix multiplications due to the enforced truncation.

6.3 Representing the conduction manifold

So far, $\mathbf{P}^{\{1\}}$ has been constrained to only contain transitions into low energy Kohn–Sham states represented by $\mathbf{P}^{\{c\}}$ for which the conduction NGWFs $\{\chi_\alpha\}$ are explicitly optimised. This has the advantage of ensuring that the resulting excitation is well represented by the localised support functions. The optimisation of $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ for their respective manifolds also has the advantage that the representation of $\mathbf{P}^{\{1\}}$ is much more compact than it would be in other localised basis set approaches making use of numerical atomic orbitals or Gaussians. Furthermore, it allows for $\{\phi_\alpha\}$ to be very strictly localised, while the generally more delocalised Kohn–Sham states of the conduction manifold can be represented by slightly less strictly localised functions $\{\chi_\beta\}$, making the representation very computationally efficient.

However, even low energy excitations often show a relatively slow convergence with the number of unoccupied states in the conduction manifold explicitly included in the calculation and some form of representing those states becomes important for convergence. While it is always possible to optimise $\{\chi_\beta\}$ for a larger subset of the conduction manifold, this is not always desirable in practice, as higher energy conduction states normally require increasingly delocalised NGWFs to be represented, resulting in a loss of sparsity and thus computational efficiency. In methods that make use of a single set of localised orbitals to represent both the occupied and unoccupied space, this issue is often avoided by including the entire unoccupied subspace representable by the basis set into the calculation via a projector (see, for example [117, 123, 124]). In the formalism presented here, a similar result can be reached by optimising $\{\chi_\alpha\}$ for the set of low energy unoccupied states that are considered to be the most important contributions to $\mathbf{P}^{\{1\}}$ and to include the rest of the conduction space manifold in an approximate manner. The simplest choice for a redefined conduction density matrix is

$$\mathbf{P}^{\{c\}} = (\mathbf{S}^{\chi})^{-1} \tag{6.31}$$

This choice trivially preserves the idempotency constraint placed on $\mathbf{P}^{\{c\}}$ and is a projec-

tion onto the entire space representable by $\{\chi_\alpha\}$. However, given that the set of $\{\chi_\alpha\}$ is not generally chosen to be orthogonal to $\{\phi_\beta\}$, the above choice for the conduction projector also includes valence Kohn–Sham states represented by $\{\chi_\alpha\}$. Thus using $(\mathbf{S}^\chi)$ as a projector does destroy the clear separation between occupied and unoccupied states, allowing $\mathbf{P}^{\{1\}}$ to contain unphysical transitions between valence states represented by $\{\chi_\alpha\}$ and valence states represented by $\{\phi_\beta\}$.

In order to avoid unphysical valence-valence transitions, the valence states have to be projected out of the redefined choice for $\mathbf{P}^{\{c\}}$, which can be achieved by redefining the conduction density matrix as:

$$\mathbf{P}^{\{\mathrm{c}\}} = \left((\mathbf{S}^\chi)^{-1} - (\mathbf{S}^\chi)^{-1} \mathbf{S}^{\chi\phi} \mathbf{P}^{\{\mathrm{v}\}} \left(\mathbf{S}^{\chi\phi}\right)^\dagger (\mathbf{S}^\chi)^{-1} \right). \tag{6.32}$$

where $\mathbf{S}^{\chi\phi}$ denotes the cross-overlap matrix between the two NGWF representations.

Note however, that $\mathbf{P}^{\{c\}}$ defined through (6.32) is generally not idempotent even for idempotent $\mathbf{P}^{\{v\}}$, due to the two different representations used. $\mathbf{P}^{\{c\}}$ is only idempotent if $\{\chi_\alpha\}$, apart from spanning the low energy conduction space, also spans the same space as $\{\phi_\beta\}$, which is generally not true since the two sets are optimised to span different manifolds. This can be seen by explicitly evaluating

$$\begin{aligned} \mathbf{P}^{\{c\}}\mathbf{S}^\chi\mathbf{P}^{\{c\}} &= \mathbf{S}^\chi - 2\,(\mathbf{S}^\chi)^{-1}\mathbf{S}^{\chi\phi}\mathbf{P}^{\{\mathrm{v}\}}\left(\mathbf{S}^{\chi\phi}\right)^\dagger(\mathbf{S}^\chi)^{-1} \\ &\quad + (\mathbf{S}^\chi)^{-1}\mathbf{S}^{\chi\phi}\mathbf{P}^{\{\mathrm{v}\}}\left(\mathbf{S}^{\chi\phi}\right)^\dagger(\mathbf{S}^\chi)^{-1}\mathbf{S}^{\chi\phi}\mathbf{P}^{\{\mathrm{v}\}}\left(\mathbf{S}^{\chi\phi}\right)^\dagger(\mathbf{S}^\chi)^{-1} \end{aligned} \tag{6.33}$$

which, assuming idempotency of $\mathbf{P}^{\{v\}}$, only reduces to the expression of the redefined $\mathbf{P}^{\{c\}}$ if $\left(\mathbf{S}^{\chi\phi}\right)^\dagger(\mathbf{S}^\chi)^{-1}\mathbf{S}^{\chi\phi} = \mathbf{S}^\phi$. This in turn is only true if $\{\chi_\alpha\}$ can represent the same space as the set of valence NGWFs and will thus generally not be true in a typical ONETEP calculation. However, it is often found that the redefined $\mathbf{P}^{\{c\}}$ is near-idempotent, especially if only a few conduction states are optimised that have a similar character to the valence states, causing $\{\chi_\alpha\}$ to be relatively similar to $\{\phi_\beta\}$. Thus it is often found that the penalty functional defined in (6.30) stays below a desired threshold during the entire calculation, but the degree to which $\mathbf{P}^{\{c\}}$ violates the idempotency constraint is generally system-dependent and is not known *a priori*.

The consequences of the lack of idempotency in the projector can be avoided by introducing a new set of NGWFs $\{\theta_\alpha\} = \{\chi_\beta\} \oplus \{\phi_\gamma\}$ such that $\{\theta_\alpha\}$ is the joint set of conduction and valence NGWFs. If it is used for representing the conduction space manifold only, the projector can be defined to be rigorously idempotent via

$$\mathbf{P}^{\{\mathrm{c}\}} = \left(\left(\mathbf{S}^\theta\right)^{-1} - \left(\mathbf{S}^\theta\right)^{-1} \mathbf{S}^{\theta\phi} \mathbf{P}^{\{\mathrm{v}\}} \left(\mathbf{S}^{\theta\phi}\right)^\dagger \left(\mathbf{S}^\theta\right)^{-1} \right). \tag{6.34}$$

This allows the use of the projector onto the entire unoccupied subspace representable by $\{\theta_\alpha\}$ without introducing any numerical problems due to lack of idempotency. Since the joint set of NGWFs is generally about twice as large as the set of conduction NGWFs, using the projector in conjunction with the joint NGWF set increases the computational complexity of the algorithm by a factor of two.

While the use of the joint set allows for a rigorous idempotent definition of the conduction operator via a projector onto the entire conduction manifold, there are a number of drawbacks in this choice. The increased computational complexity is one, but the main issue comes in the form of the strong linear dependence that $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ can exhibit if the conduction NGWFs are optimised for a small number of conduction states that have the character of anti-bonding states. This linear dependence can in some cases cause problems when attempting to invert the joint overlap matrix $\mathbf{S}^\theta$. An alternative to using the projector in (6.32) with the joint NGWF set is to use the projector as defined by equation (6.31) and the conduction NGWF set only but removing unwanted valence transitions from $\mathbf{H}^\chi$ when evaluating the TDDFT operator $\mathbf{q}^{\chi\phi}$. This can be achieved by simply redefining $\mathbf{q}^{\chi\phi}$ via

$$(q^{\chi\phi})^{\alpha\beta} = (P^{\{\mathrm{c}\}}H^{\chi}_{\mathrm{proj}}P^{\{1\}} - P^{\{1\}}H^{\phi}P^{\{\mathrm{v}\}})^{\alpha\beta} \tag{6.35}$$

$$+ (P^{\{\mathrm{c}\}}V^{\{1\}\chi\phi}_{\mathrm{SCF}}[\rho^{\{1\}}]P^{\{\mathrm{v}\}})^{\alpha\beta}. \tag{6.36}$$

where $\mathbf{H}^{\chi}_{\mathrm{proj}}$ is the projected Hamiltonian used in the conduction optimisation (see section 5.2.1). Since $\mathbf{H}^{\chi}_{\mathrm{proj}}$ contains all valence eigenstates, but has them shifted to high energies, these forbidden transitions are never mixed in when $\left\{\mathbf{P}_i^{\{1\}}\right\}$ is minimised for some low energy excitations. The assumption here is that the energy shift λ in the projected Hamiltonian is chosen so that it leaves the low energy conduction spectrum completely unaltered compared to the pure Hamiltonian in conduction NGWF representation. In practice, the method using the projected Hamiltonian and the conduction NGWF set and the method using the projection operator and the joint set are expected to produce slightly different results, given that one method is working with a significantly larger representation of the unoccupied space. The appropriateness of the different conduction manifold representations is analysed further in section 6.5.

6.4 The exchange-correlation kernel and spin

In the above discussion, any spin dependency on both the response density matrix and on the exchange correlation kernel has been ignored. The work presented here is limited to closed shell systems, where the spin dependence of the ground state density can be dropped. As discussed in chapter 4, in many situations the spin structure of the 2-

particle TDDFT Hamiltonian can be exploited and singlet and triplet excitations can be decoupled. The decoupled singlet and triplet excitations can then be solved for separately by considering two different response potentials to the density perturbation:

$$\begin{aligned} V_{\mathrm{SCF}}^{\{1\}}[\rho_{\mathrm{singlet}}^{\{1\}}](\mathbf{r}) &= 2\int \mathrm{d}^3r' \, \frac{\rho_{\mathrm{singlet}}^{\{1\}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &\quad + \int \mathrm{d}^3r' \left(\frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho_\uparrow(\mathbf{r})\delta\rho_\uparrow(\mathbf{r}')}\bigg|_{\rho_0} + \frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho_\uparrow(\mathbf{r})\delta\rho_\downarrow(\mathbf{r}')}\bigg|_{\rho_0} \right) \rho_{\mathrm{singlet}}^{\{1\}}(\mathbf{r}') \\ V_{\mathrm{SCF}}^{\{1\}}[\rho_{\mathrm{triplet}}^{\{1\}}](\mathbf{r}) &= \int \mathrm{d}^3r' \left(\frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho_\uparrow(\mathbf{r})\delta\rho_\uparrow(\mathbf{r}')}\bigg|_{\rho_0} - \frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho_\uparrow(\mathbf{r})\delta\rho_\downarrow(\mathbf{r}')}\bigg|_{\rho_0} \right) \rho_{\mathrm{triplet}}^{\{1\}}(\mathbf{r}') \end{aligned} \quad (6.37)$$

Here the up and down arrows again denote different spin indices and it is assumed that the ground state density of the system is spin-degenerate.

For the purpose of this work, only adiabatic semi-local density kernels are considered and thus any memory effects or frequency dependence in f_{xc} are ignored, as well as any long range effects. In this case, the expression for the response potential of the system can be significantly simplified, with

$$\begin{aligned} V_{\mathrm{SCF}}^{\{1\}}[\rho_{\mathrm{singlet}}^{\{1\}}](\mathbf{r}) &= 2\int \mathrm{d}^3r' \, \frac{\rho_{\mathrm{singlet}}^{\{1\}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &\quad + \left(\frac{\partial V_{\mathrm{xc}}(\mathbf{r})}{\partial\rho(\mathbf{r})}\bigg|_{\rho_0} + \frac{\partial V_{\mathrm{c}}(\mathbf{r})}{\partial\rho(\mathbf{r})}\bigg|_{\rho_0} \right) \rho_{\mathrm{singlet}}^{\{1\}}(\mathbf{r}) \qquad (6.38) \\ V_{\mathrm{SCF}}^{\{1\}}[\rho_{\mathrm{triplet}}^{\{1\}}](\mathbf{r}) &= \left(\frac{\partial V_{\mathrm{xc}}(\mathbf{r})}{\partial\rho(\mathbf{r})}\bigg|_{\rho_0} - \frac{\partial V_{\mathrm{c}}(\mathbf{r})}{\partial\rho(\mathbf{r})}\bigg|_{\rho_0} \right) \rho_{\mathrm{triplet}}^{\{1\}}(\mathbf{r}) \qquad (6.39) \end{aligned}$$

for the triplet and singlet states respectively, where $V_{\mathrm{c}}(\mathbf{r})$ denotes the correlation potential only. Thus all that is required in the case of a spin degenerate ground state and a decoupled treatment of singlets and triplets in the excited states is the second derivative of the exchange and the correlation energy with respect to the density, evaluated at the ground state density. For any class of semi-local exchange-correlation functional discussed in this work, this second derivative can be easily found as an analytic expression from the exchange-correlation energy. However, some care has to be taken with generalised gradient approximations, since the second derivative introduces higher order gradient terms that can become difficult to treat numerically on a finite real-space grid [117]. An alternative comes in the form of evaluating the exchange-correlation part of the response potential

using a finite-differences technique[31], where one can write

$$V_{\mathrm{SCF}}^{\{1\}}[\rho_{\mathrm{triplet}}^{\{1\}}](\mathbf{r}) \approx \frac{V_{\mathrm{xc}}[\rho^{\{0\}} + \epsilon\rho_{\mathrm{triplet}}^{\{1\}}](\mathbf{r}) - V_{\mathrm{xc}}[\rho^{\{0\}} - \epsilon\rho_{\mathrm{triplet}}^{\{1\}}](\mathbf{r})}{2\epsilon} - \frac{V_{\mathrm{c}}[\rho^{\{0\}} + \epsilon\rho_{\mathrm{triplet}}^{\{1\}}](\mathbf{r}) - V_{\mathrm{c}}[\rho^{\{0\}} - \epsilon\rho_{\mathrm{triplet}}^{\{1\}}](\mathbf{r})}{2\epsilon} \tag{6.40}$$

for a chosen ϵ. Numberical tests have shown to yield good agreements with the analytical results for $\epsilon = 10^{-4}$, in good agreement with [117]. For the remainder of this work, any time an adiabatic LDA functional is used, the second derivative of the exchange correlation energy is explicitly evaluated using the analytic expression, while for any other functional, the finite difference technique is used. While the finite-difference evaluation does produce a computational overhead, since the exchange correlation potential has to be recalculated repeatedly every time $\mathbf{q}^{\chi\phi}$ is evaluated, this overhead is generally quite small, at least where GGA functionals are involved. For van der Waals type functionals, the overhead is found to be more significant, making an analytic evaluation of the second derivative desirable, however since this work does not make excessive use of van der Waals type functionals, the finite-difference implementation is deemed to be sufficient.

While this work does not make use of any hybrid functionals, it should be pointed out that they can be introduced into the linear-scaling linear-response formalism in a very straightforward manner. For hybrid functionals $\mathbf{V}_{\mathrm{SCF}}^{\{1\}}$ can be split into $\mathbf{V}_{\mathrm{SCF}}^{\{1\}\mathrm{loc}}$ containing the local part of the functional and $\mathbf{V}_{\mathrm{SCF}}^{\{1\}\mathrm{HF}}$ containing the fraction of exact exchange. $\mathbf{V}_{\mathrm{SCF}}^{\{1\}\mathrm{loc}}$ can be evaluated trivially in linear-scaling effort, while the expression for $\mathbf{V}_{\mathrm{SCF}}^{\{1\}\mathrm{HF}}$ reduces to

$$\left(V_{\mathrm{SCF}}^{\{1\}\mathrm{HF}}\right)_{\alpha\gamma} = -2c_{\mathrm{HF}} \times P^{\{1\}\beta\delta} \int\int \frac{\chi_\alpha(\mathbf{r})\phi_\gamma(\mathbf{r}')\chi_\beta(\mathbf{r})\phi_\delta(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \mathrm{d}^3r\,\mathrm{d}^3r' \tag{6.41}$$

where c_{HF} denotes the fraction of Hartree-Fock exchange [99]. Note that (6.41) is simply the gradient of the Hartree-Fock exchange energy for a density $\rho^{\{1\}}(\mathbf{r})$, a term that is very similar to terms that have to be evaluated in ground state DFT using hybrid functionals. It has been long known that in ground-state DFT, the analogous gradient of the Hartree-Fock energy with respect to the ground-state density can be evaluated in linear-scaling effort [126]. For ground state DFT, such an implementation is already available in the `ONETEP` code [127]. However, it has not yet been extended to treat two different sets of

[31] Here, it should be noted that exact arbitrary-order derivatives of the exchange-correlation energy can also be obtained via algorithmic differentiation [125]. However, for the purpose of this work, a finite differences approach is used instead.

NGWFs and while all ingredients for linear-scaling TDDFT based on hybrid functionals are available in the code, the implementation will be left to future work.

6.5 Benchmark tests

In this section a number of benchmark tests are performed on small and medium sized systems to both assess the accuracy of the TDDFT approach presented here compared to conventional TDDFT approaches and to demonstrate the scaling of the algorithm. All calculations shown here are carried out using an ALDA exchange-correlation kernel based on the Perdew-Zunger exchange-correlation energy [33]. Norm-conserving pseudopotentials are used throughout this chapter. Unless specified otherwise, a minimal NGWF set of four NGWFs per carbon, oxygen, nitrogen, magnesium, gallium and arsenide atom and one NGWF per hydrogen is chosen in each of the calculations, both for the valence NGWFs $\{\phi_\alpha\}$ and the conduction NGWFs $\{\chi_\beta\}$. Furthermore, the conduction space in the TDDFT calculations is represented through a projection onto the entire unoccupied subspace and all involved density matrices are fully dense unless specified otherwise.

6.5.1 Pentacene

The first benchmark test is performed on pentacene ($C_{22}H_{14}$) because its moderate size means that it can be conveniently studied using conventional approaches to linear-response TDDFT. First, the importance of representing the unoccupied subspace correctly is established and the effects of some of the choices presented in section 6.3 are discussed. In order to do so the calculations are compared to benchmark results obtained with the NWChem quantum chemistry package [128].

For the ONETEP benchmark results, a simulation cell size of $40 \times 49 \times 40\, a_0^3$ and a kinetic energy cutoff of 750 eV are chosen throughout, while the NGWF radius is chosen to be $10\, a_0$ for $\{\phi_\alpha\}$ and $15\, a_0$ for $\{\chi_\beta\}$ unless specified otherwise. The atomic positions of the pentacene molecule are optimised at the LDA level using the above NGWF radius and kinetic energy cutoff and the same atomic positions are used in the NWChem calculations. For the NWChem reference calculations, an aug-cc-pVTZ gaussian basis set [129] is used. Core electrons are explicitly included in the ground state calculation but are removed from the active space of possible transitions in the TDDFT calculation to guarantee a good comparability with the ONETEP results.

The first test can be taken as a more detailed analysis on how the representation of the unoccupied subspace as discussed in section 6.3 influences the excitation energies obtained with the ONETEP method. For this purpose, the 10 lowest conduction states are optimised explicitly and four different cases are considered: case **A** limits the conduction

	ONETEP(**A**)	ONETEP(**B**)	ONETEP(**C**)	ONETEP(**D**)	NWChem
1	2.055(0.047)	1.903 (0.041)	1.883 (0.050)	1.885 (0.047)	1.844 (0.044)
2	2.430	2.429	2.416	2.420	2.408
3	3.003	2.996	2.961	2.986	2.961
4	3.362	3.191	3.143	3.155	3.115
5	3.434	3.434	3.419	3.422	3.412
6	3.988(0.032)	3.898(0.030)	3.852(0.034)	3.865(0.033)	3.839(0.030)
7	3.941	3.933	3.918	3.919	3.908
8	4.026	4.020	4.003	4.010	4.002
9	4.086(0.015)	4.059 (0.010)	4.029 (0.011)	4.042(0.013)	4.029(0.012)
10	4.168	4.168	4.162	4.160	4.159
⋮	⋮	⋮	⋮	⋮	⋮
(d)	4.252	4.254	4.251	4.246	4.246
(b)	-	4.171 (2.89)	4.311 (2.58)	4.278(3.85)	4.270(3.88)

Table 6.1: Results for the ten lowest singlet states of pentacene, as calculated using ONETEP in comparison with results generated by NWChem. The first three columns correspond to ONETEP calculations using different representations of the unoccupied subspace, where **A** denotes a conduction space limited to the 10 lowest optimised states, **B** $(\mathbf{S}^{\chi})^{-1}$ as a projector and $\mathbf{H}^{\chi}_{\mathrm{proj}}$, **C** uses the conduction NGWFs and the non-idempotent projector and **D** the joint NGWF set and the idempotent projector onto the entire unoccupied space. The NWChem calculations are performed using an aug-cc-pVTZ basis. Energies are given in eV, oscillator strengths in brackets.

space manifold to the 10 states only. Case **B** uses $\mathbf{S}^{\{\chi\}}$ as a projector onto the unoccupied subspace and $\mathbf{H}^{\chi}_{\mathrm{proj}}$ as the conduction Hamiltonian, where the shift parameter λ is fixed to 20 Ha. In case **C** the non-idempotent projector onto the entire unoccupied space as represented by the conduction NGWF set only (see equation (6.32)) is used, while in case **D** the joint NGWF set is used to represent the conduction space, allowing for the projection operator to be rigorously idempotent (see equation (6.34)).

The results of these calculations for the first ten singlet excitations of pentacene, as well as two selected higher energy states (one dark and one bright, labelled (d) and (b) respectively) can be found in table 6.1. The character of the excited states obtained has been analysed and compared to the NWChem results by breaking down the excitations into individual Kohn-Sham transitions. Where necessary, the ONETEP excited states have been reordered such that the same order as in the NWChem results is obtained.

Note that there is significant discrepancy between the results obtained if the conduction subspace is limited to only the ten lowest conduction states and all the other results for a number of excited states. While most of the dark states are reasonably well represented, the states with significant oscillator strength, as well as the 4th excitation, are seriously overestimated in energy, with discrepancies of up to 0.15 eV. The reason why some of the dark states are well represented by truncating the allowed unoccupied subspace is that these states show only minimal mixing of Kohn–Sham states and are well represented by a single low energy Kohn–Sham transition.

This phenomenon is easily understood. In general, transitions that have a vanishing linear response term $\mathbf{V}^{\{1\}}_{\mathrm{SCF}}$ associated with them tend to be pure low energy Kohn–Sham transitions, while large response terms $\mathbf{V}^{\{1\}}_{\mathrm{SCF}}$ imply that the excitation energy can be lowered by mixing in higher energy Kohn–Sham transitions that minimise the system response. Large off-diagonal response terms are often associated with high oscillator strengths, while transitions with vanishing system response tend to be pure Kohn–Sham transtions with small oscillator strengths. Thus limiting the unoccupied subspace to a small energy window naturally causes a number of dark states to be very well represented but produces significant errors for other states.

Comparing the other three representations of the unoccupied subspace with the benchmark results from NWChem, very good agreement is found, both between the individual ONETEP results and with the benchmark calculation. Average discrepancies per excitation between the ONETEP methods for the first ten states is of the order of 0.01 eV, while the average errors compared to the NWChem results are slightly larger. The main discrepancies occur in the first and 4th state, precisely the excitation that is very badly represented by truncating the conduction subspace to 10 states and therefore a state that contain significant mixing with higher energy Kohn–Sham states. Naturally, this state is

	ONETEP (**1H**)	ONETEP (**2H**)	ONETEP (**5H**)	NWChem
1	1.883 (0.050)	1.855 (0.049)	1.839 (0.050)	1.844 (0.044)
2	2.416	2.402	2.405	2.408
3	2.961	2.942	2.945	2.961
4	3.143	3.121	3.103	3.115
5	3.419	3.405	3.409	3.412
6	3.852(0.034)	3.831(0.035)	3.821(0.035)	3.839(0.030)
7	3.918	3.900	3.903	3.908
8	4.003	4.000	3.996	4.002
9	4.029 (0.011)	4.032 (0.013)	4.006(0.013)	4.029(0.012)
10	4.162	4.106	4.101	4.159
⋮	⋮	⋮	⋮	⋮
(d)	4.251	4.216	4.211	4.246
(b)	4.311(2.58)	4.281(3.87)	4.239(3.92)	4.270(3.88)

Table 6.2: Results for the excited states of pentacene, as calculated using ONETEP with the non-idempotent projection onto the entire unoccupied subspace and the conduction NGWF representation, in comparison with results generated by NWChem. Results are shown for the 10 lowest excitations, as well as two selected higher energy states, one dark and one bright (labelled (d) and (b) respectively). The first three columns correspond to ONETEP calculations using three different conduction NGWF representation, with 1, 2 and 5 NGWFs per H atom respectively. The NWChem calculations are performed using an aug-cc-pVTZ basis. Energies are given in eV, oscillator strengths in brackets. This table is reproduced from [99].

not perfectly represented in ONETEP by only optimising $\{\chi_\alpha\}$ for the 10 lowest conduction states.

The main discrepancy in the ONETEP results is obtained for the bright high energy excited state, for which both results sets using $\{\chi_\alpha\}$ to represent the unoccupied states fail to give a decent description of the excitation energy and the oscillator strength but for which the results set using the joint NGWF representation is in excellent agreement with the NWChem results. In general, the joint representation of the conduction space in combination with the idempotent projector onto the entire unoccupied subspace shows the most consistent agreement with the NWChem results over the wide range of excitations tested here. The main premise made in the previous sections holds: while it is evidently very important to have some form of representation of the entire unoccupied manifold in order to obtain well-converged excitation energies, good results are achieved for optimising the NGWFs for only a small number of conduction states explicitly and representing the rest of the unoccupied space approximately.

It should be noted that the ONETEP results are achieved using only a minimal number of NGWFs. The effective dimensions of the two-particle TDDFT Hamiltonian in NGWF representation are thus 10404×10404 if the conduction NGWF set is used to represent

the unoccupied subspace and 20808×20808 if the joint set is used. In contrast, the aug-cc-pVTZ basis set in NWChem generates an effective two-particle Hamiltonian of dimensions of approximately $(1.8 \times 10^6) \times (1.8 \times 10^6)$, which is significantly larger in size, making the ONETEP representation very compact and efficient.

It can be seen that the minimal NGWF representation can achieve very good results if the unoccupied subspace is introduced into the calculation in some approximate form. However, it is also demonstrated that some states in the low energy spectrum of pentacene have a significant amount of high energy conduction states mixed in that might be only poorly representable by the minimal set of NGWFs chosen. In order to test for this, the calculation performed with the conduction NGWF set only and with the nonidempotent projector onto the unoccupied space is taken and repeated for different sets of conduction NGWFs. Specifically, the conduction NGWFs on the hydrogen atoms are enhanced by choosing two and five NGWFs per hydrogen instead of the minimal representation of a single NGWF. It is to be assumed that the additional variational freedom on the hydrogen atoms leads to a better representation of some very delocalised unoccupied states. When optimising $\{\chi_\alpha\}$ the additional variational freedom allows for explicitly converging the first 14 conduction states for the enhanced NGWF representations instead of only the first 10 for the minimal set.

The results of this benchmark calculation as compared to the NWChem results are summarised in table 6.2. Note that the enhanced NGWF representations indeed systematically lower the energy of the excitations that have a significant amount of high energy conduction states mixed in, leaving other states virtually unaltered. Enhancing the NGWF representation leads to a significant improvement of the character of the bright state, with the oscillator strength in line with the NWChem results for both 2 and 5 NGWFs on the H atoms. In general, an improved agreement with the benchmark results is found for the two enhanced representations, suggesting that representing the very diffuse high energy conduction states correctly becomes important in a number of low energy states in pentacene. It should be pointed out that even the enhanced representations chosen here are significantly smaller in size than the NWChem representation, with the largest representation corresponding to a two-particle Hamiltonian of dimensions 24336×24336. It can be concluded that the compact ONETEP NGWF representation of the conduction-valence product space can yield results that are in excellent agreement with Gaussian representations that are larger by a factor of 100.

It should be pointed out that while NWChem is treated as a benchmark in this work, it also uses a localised orbital representation and is thus expected to experience similar drawbacks to ONETEP when converging excitations into very delocalised conduction states. In many respects, the NGWF approach is at an advantage here, since it is always possible

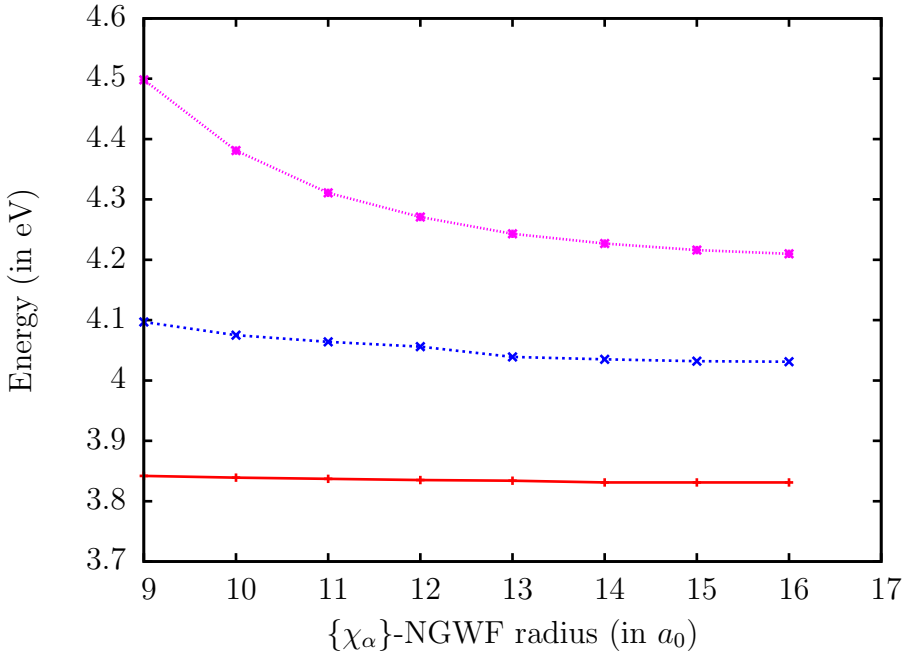


Figure 6.1: Convergence of three selected excitation energies of pentacene with conduction NGWF radius (adapted from [99]).

to increase the NGWF localisation region and optimise higher energy unbound conduction states to get a good representation of delocalised excitations. Note for example that for the states labelled (b) and (d) which have significant contributions from transitions into delocalised conduction states, the enhanced NGWF representations seem to converge to significantly lower excitation energies than the NWChem benchmark.

As a final benchmark test on the pentacene molecule, the dependence of the excited state energies on the NGWF localisation radius of $\{\chi_\alpha\}$ is established. The minimal NGWF representation and the non-idempotent projector unto the unoccupied space is chosen again for this purpose and the NGWFs are optimised for the 10 lowest conduction states. The results for three selected excited states can be found in figure 6.1, where the states shown correspond to states labelled 6, 9 and (d) in table 6.2 respectively.

As it can be seen from the data, the three states show very different convergence behaviour. While the 6th singlet state is already well converged for an NGWF radius of $9\,a_0$, the dark state shows a much slower convergence behaviour, requiring a radius of $15\,a_0$ to reach convergence. The different convergence behaviours can be explained by breaking down the excitations into individual Kohn–Sham transitions, where it is found that the dark state is composed of 99 % of a transition from the HOMO to the 9th unoccupied state. The 9th unoccupied Kohn–Sham state is only very lightly bound and diffuse and thus naturally requires a larger NGWF radius to be represented correctly. However, while different excited states show different convergence behaviours depending on the

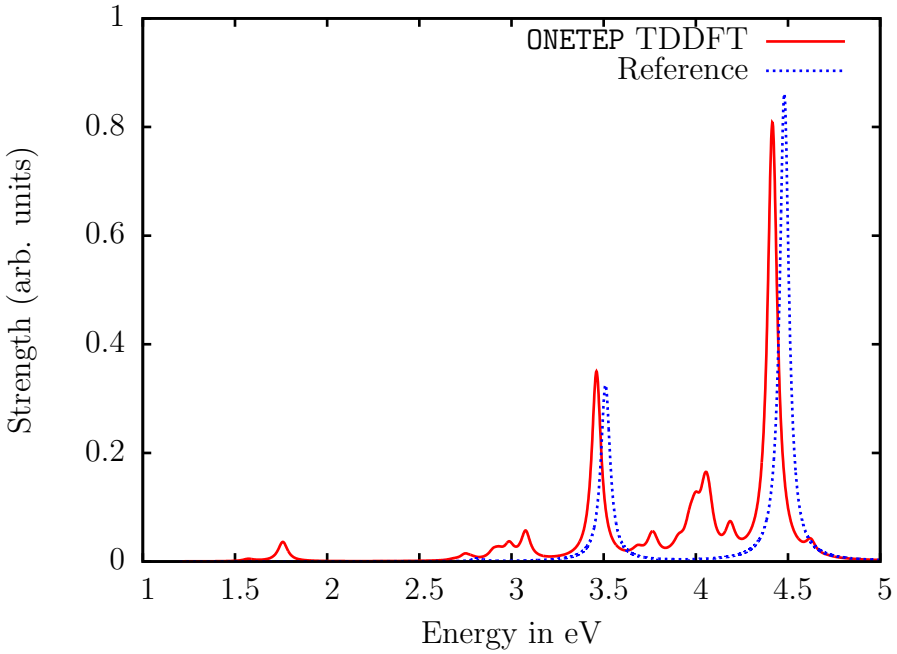


Figure 6.2: Low energy absorption spectrum of C_{60} calculated with ONETEP as compared to results of the three main transitions obtained from [130](figure has been adapted from [99]). A Lorentzian broadening of 0.03 eV has been used on both the reference results and the ONETEP results and a constant shift in absorption strength has been applied to the reference results in order to make them comparable.

character of the Kohn–Sham states involved in the transition, the results demonstrate that the NGWF method is very capable of representing even diffuse excited states and that increasing the localisation radius leads to a systematic convergence of excitation energies.

6.5.2 Buckminsterfullerene

The next test system considered in this work is the C_{60} molecule that has been studied in great detail by conventional TDDFT implementations. Since the low energy spectrum of C_{60} is dominated by a few very strong peaks and a large number of dark states, it provides a good test system to demonstrate the scaling of the method with the number of individual excitations converged. The ONETEP results are obtained using a box size of $37.8 \times 37.8 \times 37.8\, a_0^3$ and a kinetic energy cutoff of 800 eV. The NGWF radii of $13\, a_0$ and $8\, a_0$ for the conduction and valence representations respectively are found to yield well converged results while the conduction NGWFs are optimised for the 30 lowest unoccupied states. The conduction space manifold is represented by $\{\chi_\alpha\}$ and the non-idempotent projector is used for the purpose of this calculation.

The low energy spectrum of C_{60} as generated from the lowest 150 excitations can

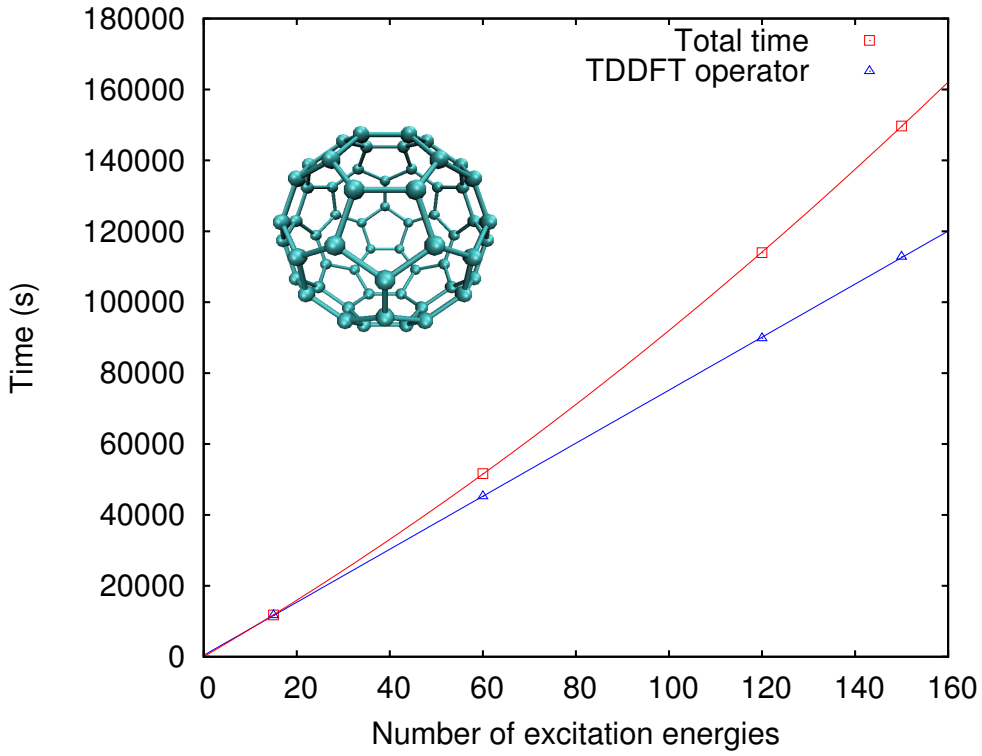


Figure 6.3: Time taken during a full calculation of applying the TDDFT operator when calculating the 150 lowest excitations of C_{60}. The blue and red lines are linear and quadratic fits to the data points respectively. The figure is reproduced from [99].

be found in figure 6.2, where it is compared to the three main transitions taken from a reference calculation found in [130] that was obtained with a 6-31G+s[32] Gaussian basis set and the BP86 GGA functional. The two main peaks and their relative oscillator strengths are found to be in good agreement with the reference calculations, with the `ONETEP` results being slightly lower in energy. However, as noted by the authors of the reference calculation, the Gaussian basis set used is relatively small and estimated errors to excitation energies are around 0.1 eV.

In figure 6.3, the scaling of the TDDFT algorithm with respect to the number of excited states converged is demonstrated. As can be seen, the time taken to apply the TDDFT operator to a batch of trial transition vectors scales linearly with the number of transition vectors. However, the total time taken by the algorithm to converge the lowest N_ω excited states shows a scaling that is quadratic in character. This confirms the analysis in section 6.1.4, in that the $\mathcal{O}(N_\omega^3)$ asymptotic scaling of the subspace diagonalisation is completely insignificant even when $N_\omega = 150$ compared to the $\mathcal{O}(N_\omega^2)$ scaling of the Gram-Schmidt orthonormalisation that has to be carried out at the end of each conjugate gradient iteration.

[32]This denotes a 6-31G basis set that is augmented by diffuse s-functions. A more detailed description on how the basis set was created can be found in [130].

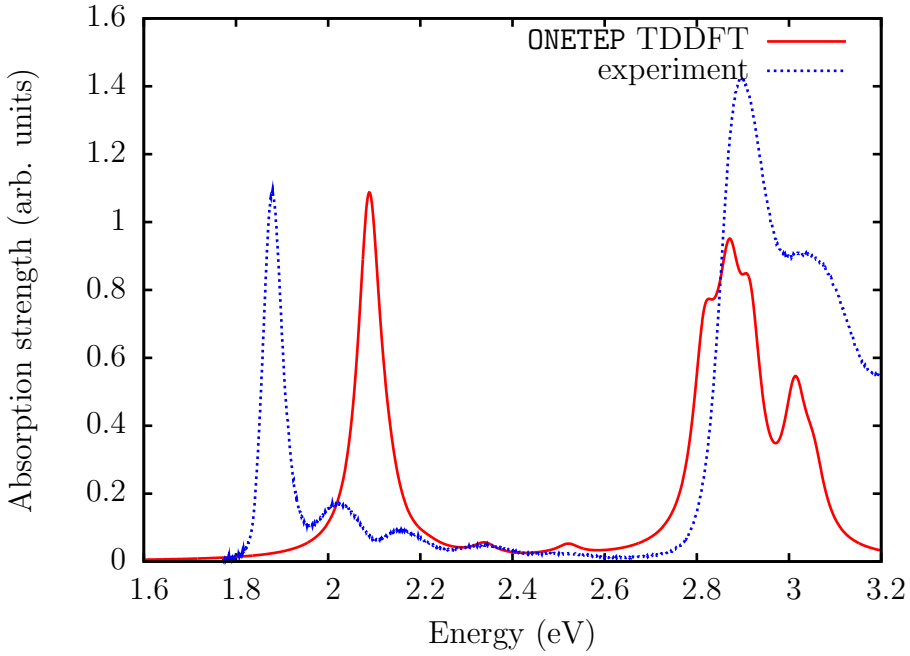


Figure 6.4: Excitation spectra generated from the 12 lowest singlet excitations of chlorophyll a in vacuum, compared to an experimental spectrum in diethyl ether [131](figure adapted from [99]). A Lorentzian broadening of 0.03 eV was used for the TDDFT results.

6.5.3 Chlorophyll

Chlorophyll a ($MgN_4O_5C_{55}H_{72}$) can be considered an ideal application for the method developed here. Although it is too small to take advantage of the linear-scaling capabilities of the ONETEP method, it has a low energy spectrum that is completely determined by a few transitions with large oscillator strength, while containing very few dark states. Thus in contrast to C_{60}, where a large number of excited states have to be converged in order to obtain a meaningful low energy spectrum, Chlorophyll a is much more suited to the linear-response TDDFT approach considered here.

Calculations are performed using a kinetic energy cutoff of 800 eV and a localisation radius of $12.0\,a_0$ and $8.0\,a_0$ is used for the $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ NGWF sets respectively. The lowest 15 conduction states are explicitly optimised and the non-idempotent projector onto the entire unoccupied subspace is used.

Figure 6.4 shows the low energy spectrum of chlorophyll a generated from the 12 lowest singlet states in comparison to experimental results of chlorophyll in diethyl ether [131]. Note that the general shape of the spectrum, showing a sharp peak due to a single transition in the low energy range and an extended feature at around 3.0 eV, is well reproduced by the TDDFT calculation. In general, the TDDFT spectrum shows a blue shift compared to the experimental results, in good agreement to published TDDFT results, both within the Tamm–Dancoff approximation and a planewave basis and full

TDDFT and a localised orbital representation [123, 124, 132].

Comparing individual excitations, the position of the first peak in the spectrum is found to be at 2.057 eV with an oscillator strength of 0.208, in good agreement to the 2.00 eV and oscillator strength of 0.182 obtained by Sundholm [132] for an SP(V) Gaussian basis set. The onset of the extended feature is caused by an excitation at 2.803 eV and an oscillator strength of 0.133, while Sundholm reports an energy of 2.75 eV and an oscillator strength of 0.182 for the same transition. The `ONETEP` results are systematically higher in energy and while in Sundholm's result the two main spectral features in the low energy range are caused by excitations with identical oscillator strength, the `ONETEP` results show some shift of spectral weight to the lowest singlet excitation of the system. Note however, that the differences between the results presented here and the published results in [132] are likely to originate from the fact that this work makes use of the Tamm–Dancoff approximation, which is known to raise the excitation energy of bright states by as much as tenths of eVs in some cases and cause changes in spectral weightings[33].

In summary, the obtained results for chlorophyll *a* that are in reasonably good agreement with the ones published in the literature. It can be concluded that systems like chlorophyll, with low energy spectra that are completely dominated by a few strong transition and that do not contain a large number of dark states, are an ideal application for the method developed in this work, allowing for very accurate and efficient calculations.

6.5.4 Carbon nanotube

To demonstrate the capabilities of the algorithm to perform TDDFT calculations in linear-scaling effort, a number of benchmark calculations on a (10,0) single-walled carbon nanotube (CNT) are carried out in periodic boundary conditions. Different supercell sizes containing 640, 920, 1240, 1600 and 1920 atoms are chosen, corresponding to segments of 127, 193, 257, 321 and 386 a_0 in length.

All calculations are performed with a 700 eV kinetic energy cutoff and $8\,a_0$ and $12\,a_0$ radii for the $\{\chi_\alpha\}$- and $\{\phi_\beta\}$-NGWF representations respectively. The conduction NGWFs are optimised for all bound states and no projector onto the entire unoccupied subspace is used. Thus, the size of the allowed unoccupied subspace for the TDDFT transitions does increase linearly with system size. For the largest CNT supercell containing 1920 atoms, the dimensions of the active canonical Kohn–Sham transition space are $(1.84 \times 10^6) \times (1.84 \times 10^6)$, while the 2-particle Hamiltonian in effective conduction-valence NGWF representation has a size of $(5.90 \times 10^7) \times (5.90 \times 10^7)$, prohibitively large for any non-iterative treatment of the eigenvalue problem. In order to achieve a linear scaling computational effort with system size, the conduction and valence density matrices

[33] For an example of this effect as observed with polycyclic aromatic hydrocarbons, see [133]

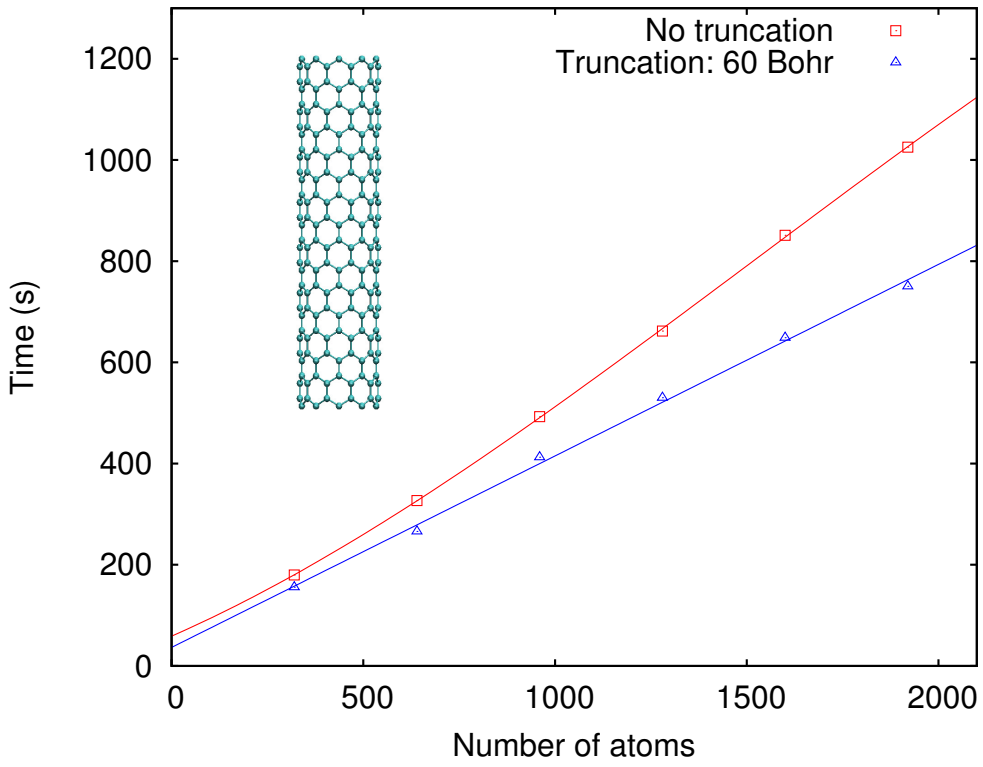


Figure 6.5: Time taken for a single conjugate gradient iteration for a (10,0) carbon nanotube with different numbers of atoms in the system for fully dense density matrices and for truncated density matrices with a truncation radius of 60 a_0. The blue line represents a linear fit while the red line is a cubic fit to the data points. The calculations were performed using 6 Intel Ivy Bridge nodes with 12 cores each. This figure is reproduced from [99].

are truncated using a spherical cutoff radius of $R_{\text{cut}} = 35\,a_0$.

Since the calculations are performed in periodic boundary conditions, all supercell sizes correspond to the case of an infinite CNT. Due to the infinite system size, the lowest excitation of the system corresponds to an excited state delocalised along the entire length of the rod. This causes the response density and thus the system response $\mathbf{V}_{\text{SCF}}^{\{1\}}$ to vanish and the lowest excitation energy to reduce to the Kohn–Sham gap for any semi-local exchange-correlation functional, which is a well known failure of TDDFT in periodic systems (see the discussion in section 4.2.6). Therefore, the results presented here are to be taken as a demonstration of linear-scaling capabilities only, while linear-scaling calculations of more physically meaningful systems will be demonstrated in later chapters.

Figure 6.5 shows the time taken for performing a single conjugate gradient iteration of a single excited state over the range of supercell sizes discussed above. Two different timings are shown: one for the case of a fully dense response density matrix and one where a moderate truncation of $R_{\text{cut}} = 60\,a_0$ is applied. Note that as soon as the truncation is

applied, the time taken for a single conjugate gradient iteration scales fully linearly with system size. If $\mathbf{P}^{\{1\}}$ is not truncated, the asymptotic scaling goes back to the usual $\mathcal{O}(N^3)$. However, it can be seen that even if no truncation is applied, the algorithm presented here exhibits near linear scaling up to the largest system sizes considered. The reason for this can be seen in the very compact NGWF representation chosen in ONETEP, meaning that the fully dense matrix-matrix multiplications required in the conjugate gradient algorithm only become a significant contribution to the total calculation time for system sizes beyond 2000 atoms. Thus while it is necessary to truncate all density matrices to enable the calculation of very large system sizes in ONETEP, this is not strictly necessary for a range of medium-sized systems that are already inaccessible by conventional TDDFT methods that use a less compact representation.

6.6 Convergence properties and preconditioning

In order to analyse the convergence properties of the conjugate gradient algorithm described here, it is important to consider some properties of the original 2-particle Hamiltonian matrix $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$ that is solved for its lowest eigenvalues. It can be assumed, as a first order approximation, that $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$ is diagonally dominant in canonical Kohn–Sham space, such that, in product NGWF space $\mathbf{q}^{\chi\phi} \approx \mathbf{q}^{\chi\phi}_{\mathrm{diag}}$ with

$$\mathbf{q}^{\chi\phi}_{\mathrm{diag}} = \mathbf{P}^{\{\mathrm{c}\}}\mathbf{H}^{\chi}\mathbf{P}^{\{1\}} - \mathbf{P}^{\{1\}}\mathbf{H}^{\phi}\mathbf{P}^{\{\mathrm{v}\}}. \tag{6.42}$$

The approximation used here is that the self-consistent response of the system to the perturbation caused by the neutral excitation is generally small, which has to be the case in order for linear-response theory to be applicable in the first place. Approximating $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$ as diagonally dominant, one can find an approximate expression for the condition number of the two-particle Hamiltonian matrix. While there are several different ways of stating the condition number of a matrix, the most convenient way in this scenario is via the eigenvalues of $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$, such that

$$\kappa(\mathbf{H}^{[2p]}_{\mathrm{TDA}}) = \left| \frac{\omega_{\mathrm{max}}}{\omega_{\mathrm{min}}} \right| \tag{6.43}$$

with ω_{max} and ω_{min} being the maximum and minimum eigenvalues of $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$. Approximating the condition number of the diagonally dominant $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$ with that of the diagonal matrix of Kohn–Sham eigenvalue differences $\mathbf{H}^{[2p]}_{\mathrm{diag}}$, it is found that

$$\kappa(\mathbf{H}^{[2p]}_{\mathrm{TDA}}) \approx \kappa(\mathbf{H}^{[2p]}_{\mathrm{diag}}) = \left| \frac{\epsilon^{\mathrm{KS}}_{\mathrm{max}} - \epsilon^{\mathrm{KS}}_{\mathrm{min}}}{\epsilon^{\mathrm{KS}}_{\mathrm{LUMO}} - \epsilon^{\mathrm{KS}}_{\mathrm{HOMO}}} \right| = \left| \frac{E_{\mathrm{window}}}{E_{\mathrm{gap}}} \right| \tag{6.44}$$

where E_{gap} denotes the Kohn–Sham bandgap and E_{window} denotes the size of the active energy window between the lowest valence state and highest conduction state explicitly included in the calculation. From this, two conclusions can be drawn:

- The matrix $\mathbf{H}_{\mathrm{TDA}}^{[2p]}$ naturally becomes increasingly ill-conditioned as the Kohn–Sham bandgap of the system decreases.
- Generally, $E_{\mathrm{gap}} \ll E_{\mathrm{window}}$ and thus $\mathbf{H}_{\mathrm{TDA}}^{[2p]}$ is ill-conditioned even for non-vanishing Kohn–Sham bandgaps.

The ill-contitioning of the eigenvalue problem causes the conjugate gradient to show a poor convergence behaviour. This problem can be avoided by using an appropriate preconditioner. Assuming again that the diagonal part of the Kohn–Sham eigenvalue differences dominates over the coupling term due to the system response, $\mathbf{H}_{\mathrm{diag}}^{[2p]}$ can be assumed to be a very effective preconditioner, such that one can solve the left-preconditioned eigenvalue problem

$$\left(\mathbf{H}_{\mathrm{diag}}^{[2p]}\right)^{-1} \mathbf{H}_{\mathrm{TDA}}^{[2p]} \mathbf{x} = \omega \left(\mathbf{H}_{\mathrm{diag}}^{[2p]}\right)^{-1} \mathbf{x} \tag{6.45}$$

where $\kappa\left(\left(\mathbf{H}_{\mathrm{diag}}^{[2p]}\right)^{-1} \mathbf{H}_{\mathrm{TDA}}^{[2p]}\right)$ is close to 1. In conduction-valence Kohn–Sham space, applying this preconditioner is very easy, since $\mathbf{H}_{\mathrm{diag}}^{[2p]}$ is diagonal in this representation and its inverse is trivially found. In NGWF space, $\mathbf{H}_{\mathrm{diag}}^{[2p]}$ is non-diagonal and the inverse cannot be found in linear-scaling effort by direct linear algebra methods. Furthermore, $\mathbf{H}_{\mathrm{diag}}^{[2p]}$ is never explicitly constructed, making the direct calculation of its inverse undesirable. Instead, one can obtain the preconditioned version $\mathbf{f}_{\mathrm{pre}}^{\chi\phi}$ of the steepest descent search direction $\mathbf{f}^{\chi\phi}$ by solving the linear system

$$\mathbf{f}^{\chi\phi} = \mathbf{P}^{\{c\}} \mathbf{H}^{\chi} \mathbf{f}_{\mathrm{pre}}^{\chi\phi} - \mathbf{f}_{\mathrm{pre}}^{\chi\phi} \mathbf{H}^{\phi} \mathbf{P}^{\{v\}} \tag{6.46}$$

for $\mathbf{f}_{\mathrm{pre}}^{\chi\phi}$, which can be easily done in linear scaling effort using a conjugate gradients algorithm.

Figure 6.6 shows the number of iterations needed to converge the four lowest singlet excitations of pentacene for different levels of preconditioner applied. Clearly, there is a significant improvement in convergence for the preconditioned system compared to the case when no preconditioner is used, in good agreement with results obtained in [98]. The preconditioned case converges in less than 1/6th of the number of iterations needed the unconditioned system. However, more importantly, the results show that it is not necessary to solve the linear system in (6.46) to numerical accuracy to obtain a significant increase in performance. Limiting the solution to the linear system to a fixed number of conjugate gradient steps corresponds to only applying the preconditioner approximately

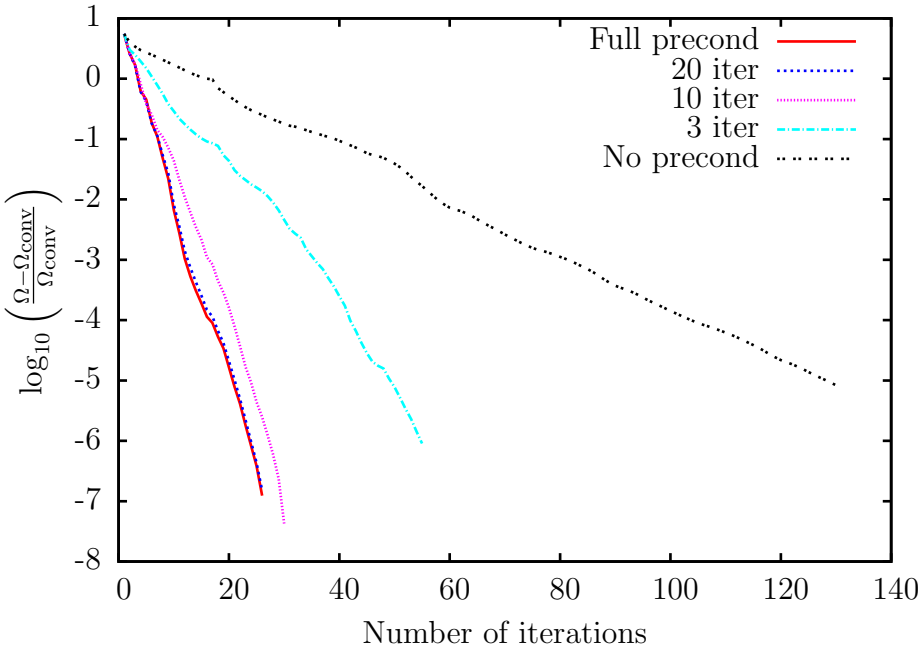


Figure 6.6: Logarithmic convergence of Ω for the four lowest states of pentacene compared to the asymptotic value Ω_{conv} for different degrees to which the preconditioner is applied. For "No Precond", no preconditioner is applied while for "Precond", the linear system of the preconditioner is solved to numerical accuracy. For the other three lines, the conjugate gradient algorithm solving the linear system of the preconditioner is limited to 20, 10 and 3 iterations respectively.

every iteration. In every case the convergence properties are significantly improved compared to the case where no preconditioning is used, even in the extreme example when only three conjugate gradient steps are performed in each iteration.

This is an important finding, as it justifies limiting the solution of the linear system to a few iterations in the case of large systems. In large systems, even if all involved matrices are sparse, performing the matrix multiplications associated with the Kohn–Sham eigenvalue difference part of $\mathbf{q}^{\chi\phi}$ can become comparable to the computational cost of evaluating $\mathbf{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}[\mathbf{P}^{\{1\}}]$. Once this happens, solving (6.46) to numerical accuracy in every single conjugate gradient iteration becomes unfeasible and a good balance has to be found between speeding up the convergence by accurately applying the preconditioner and saving computational effort by only applying it approximately[34].

[34]While there is no way to know *a priori* how to find this balance in practical calculations, it can be easily determined by performing a timed single conjugate gradient step at the beginning of the calculation. Since even a very approximate application of the preconditioner is found to double the convergence rate, choosing the number of preconditioner iterations such that they do take up less than 1/3 of the time of the entire conjugate gradient iteration is a safe approach to ensure a good performance.

6.7 Including solvation effects

In many situations one would like to calculate the spectrum of a biological molecule not in vacuum but rather in a solvent. As discussed in section 5.2.3, a computationally feasible alternative to treating solvent molecules explicitly is to introduce the solvent effects on the electron density via an implicit solvent model. The implicit solvent formalism can be straightforwardly extended to LR-TDDFT calculations. Performing a ground state and conduction state optimisation in the implicit solvent as described in section 5.2.3 yields Hamiltonians $\mathbf{H}^{\chi}$ and $\mathbf{H}^{\phi}$ and Kohn–Sham density matrices $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{v\}}$ that fully include all solvent effects. Thus the diagonal term of Kohn–Sham eigenvalue differences is consistent with the implicit solvent approach as long as the Hamiltonians and Kohn–Sham density matrices are obtained from the implict solvation calculation.

However, some care has to be taken when considering solvent effects to the self-consistent response potential $\mathbf{V}^{\chi\phi}_{\mathrm{SCF}}[\mathbf{P}^{\{1\}}](\mathbf{r})$. The ground state of the system, and thus the diagonal term of the LR-TDDFT equation, is fully screened by the static dielectric constant $\epsilon^{\infty}_{\mathrm{static}}$ of the solvent. This static dielectric constant models the rearrangement of solvent molecules to maximise the screening of the solute molecule, which occurs over a long timescale. The perturbation in the ground state density due to the excitation on the other hand occurs on a very fast timescale, on which the slow degrees of freedom of the solvent cannot react to the change in the electron density of the solute [134]. In this "non-equilibrium regime", the solvent effects are therefore modelled by an effective $\epsilon^{\infty}_{\mathrm{dynamic}}$ accounting for a screening due to the fast degrees of freedom of the solvent only.

The Hartree potential of the perturbed ground state density of the system can then be computed via:

$$\nabla \cdot \left(\epsilon_{\mathrm{dynamic}}[\rho^{\mathrm{pert}}]\nabla\Phi^{\mathrm{pert}}\right) = -4\pi\rho^{\mathrm{pert}}_{\mathrm{tot}} \tag{6.47}$$

where $\rho^{\mathrm{pert}} = \rho_0 + \rho^{\{1\}}$, the perturbed ground state density due to the excitation and $\rho^{\mathrm{pert}}_{\mathrm{tot}}$ is the sum of the perturbed density ρ^{pert} and the Gaussian-smeared density of ion cores. Here, the approximation is made that the solvation cavity for evaluating $\epsilon[\rho^{\mathrm{pert}}]$ is taken to be fixed and defined by the ground state density in vacuum. Then $V^{\chi\phi}_{\mathrm{H}}[\mathbf{P}^{\{1\}}](\mathbf{r})$ can simply be written as

$$V^{\chi\phi}_{\mathrm{H}}[\mathbf{P}^{\{1\}}](\mathbf{r}) = \Phi^{\mathrm{pert}}(\mathbf{r}) - \Phi_0(\mathbf{r}) \tag{6.48}$$

where $\Phi_0(\mathbf{r})$ is the molecular Hartree potential due to the ground state density and only has to be evaluated once at the beginning of a TDDFT calculation. The fact that the solvation cavity is forced to stay fixed during the TDDFT calculation means that the response density cannot respond self-consistently to the polarisation of the dielectric. Furthermore, the fact that the perturbed density ρ^{pert} is often more delocalised than ρ_0 due to the delocalised nature of the conduction Kohn-Sham states [113] is also ignored

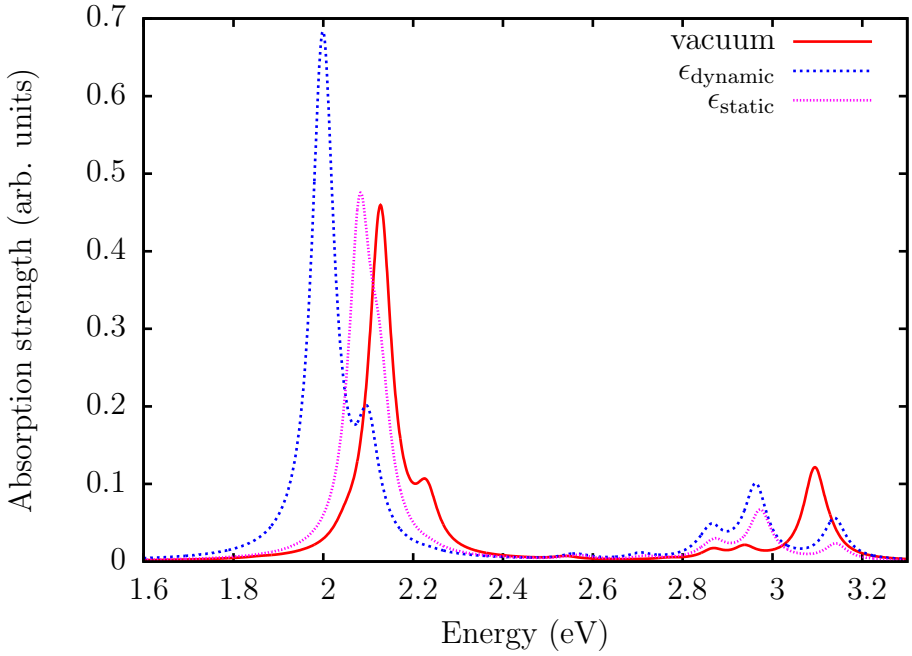


Figure 6.7: Excitation spectra generated from the 12 lowest singlet excitations of Bacteriochlorophyll in vacuum, in a solvent of $\epsilon_{\text{static}} = 80$ and in the case where dynamic solvent effects are approximated by $\epsilon_{\text{dynamic}} = 2$ and are included in $V_{\text{SCF}}^{\{1\}}$. A Lorentzian broadening of 0.03 eV is used.

in the fixed cavity approach. This can in principle be remedied by allowing the cavity to vary self-consistently, with the effect that the expression (6.48) has to be adapted to

$$V_{\text{H}}^{\chi\phi}[\mathbf{P}^{\{1\}}](\mathbf{r}) = \Phi^{\text{pert}}(\mathbf{r}) - \frac{1}{8\pi}\left(\nabla\Phi^{\text{pert}}(\mathbf{r})\right)^2 \frac{\delta\epsilon}{\delta\rho^{\text{pert}}}(\mathbf{r}) - \Phi_0(\mathbf{r}) \tag{6.49}$$

Just like in standard implicit solvent calculations the second term on the right hand side of Equation (6.49) is difficult to treat numerically and needs a very fine real space grid to be resolved correctly. For this reason, for the purpose of this work the cavity will be approximated as being fixed to the ground state cavity, which is more in line with a linear-response type treatment of the TDDFT excitation energies.

The effect of the implicit solvent model on the TDDFT excitation energies is demonstrated by considering a bacteriochlorophyll (Bchl) molecule. The atomic positions of the molecule are taken from the X-ray diffraction data [135] for the Bchl site 3 in the Fenna-Matthews-Olson complex and are kept fixed. The low energy excitation spectrum of Bchl is then calculated both in vacuum and in water (using a dielectric constant of $\epsilon_{\text{static}}^{\infty} = 80$ and $\epsilon_{\text{dynamic}}^{\infty} = 2$). Figure 6.7 shows the changes in the low energy excitation spectra in both cases. The strong low energy peak at around 2.13 eV for Bchl in vacuum corresponds to the Qy transition of the molecule. This Qy transition is shifted consider-

ably lower in energy, to around 2.00 eV, in the solvent, where it actually corresponds to the lowest excitation of the system. The oscillator strength of the Qy transition in the solvent is enhanced. The Qx state also shows an enhanced oscillator strength in the solvent and is shifted to 2.10 eV. Furthermore, the solvated system shows some reorganisation of spectral weight in the extended feature of the molecule between 2.8 and 3.2 eV.

Figure 6.7 also shows the spectrum for the case where both the valence and conduction Hamiltonians are constructed in the implicit solvent model, while the dynamic solvation effects are ignored for the Hartree term in the TDDFT equation, meaning that the solvent effects are limited to $\mathbf{H}_{\mathrm{diag}}^{[2p]}$ only. It can be seen that the solvent effect on the Kohn–Sham energies leads to a slight shift of the Qy transition to lower energy, but only has a relatively small effect on the spectral weight. The large enhancement in spectral weight of the Qy and Qx transitions is thus entirely due to dynamic screening effects not accounted for by applying the dielectric screening to the ground state of the system only.

It can be concluded that the influence of the solvent on excitation energies and oscillator strength, as well as the ordering of states is a significant factor and needs to be correctly accounted for in solvated systems. Furthermore, figure 6.7 can be seen as a demonstration that TDDFT excitations are sensitive not only to static screening due to the electrostatic environment, but also due to dynamic screening effects. It should be noted here that while the implicit solvation model accounts for a an average screening effect, it is often found that excitation energies of chromophores in solvent are quite sensitive to the treatment of the surroundings and thus including explicitly some degree of quantum mechanical description of the environment can become necessary [113]. This premise will be revisited and discussed in some detail in chapter 9.

6.8 Sparsity of the response matrix revisited: GaAs nanorod

While the computation of the lowest singlet excitation in the carbon nanotube demonstrates the linear-scaling of the method if all density matrices are truncated, it does not provide a realistic test system, as the TDDFT corrections vanish in the periodic case due to the locality of the ALDA functional. It is important to demonstrate that density matrix truncations are indeed possible for a periodic system, however, a more realistic system is needed in order to determine the effect of truncating the response density matrix. For this reason, consider a hydrogen-terminated GaAs nanorod containing 996 atoms and having a total length of $159\,a_0$. Calculations are performed using a kinetic energy cutoff of 400 eV and an NGWF radius of $12\,a_0$ is chosen for all NGWFs. No truncations are applied to the conduction and valence density matrix.

The specific system has already been studied in great detail with the `ONETEP` method

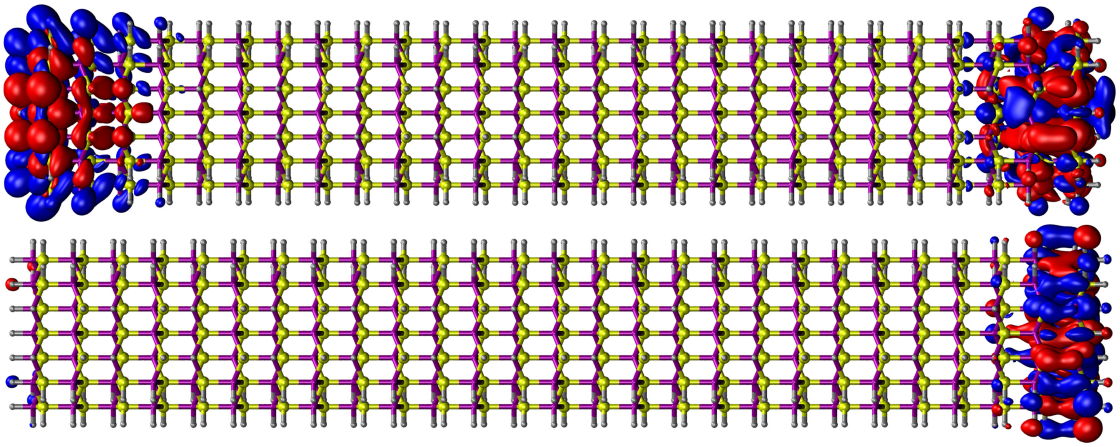


Figure 6.8: Transition density of the lowest excitation of a GaAs nanorod as found for a truncated density matrix at $75\,a_0$ (lower figure) and the full density matrix (upper figure). In this plot, H is shown in grey, As in yellow and Ga in purple. This figure has been reproduced from [99].

[136,137]. What makes the GaAs nanorod interesting from a physical point of view is that it exhibits a very strong dipole moment along its long axis. This dipole moment causes a strong electrostatic potential across the rod and forces the HOMO and LUMO to be localised on opposite ends. Given that the HOMO and LUMO are separated in space, their overlap vanishes, $\rho^{\{1\}}(\mathbf{r}) \to 0$ everywhere for a transition from HOMO to LUMO and the lowest TDDFT excitation computed with any semilocal functional approaches E_{gap}. Thus the lowest excitation of the system as computed with an ALDA functional is expected to be a spurious charge transfer state corresponding to moving an electron from one end of the rod to the other.

For a fully dense response matrix $\mathbf{P}^{\{1\}}$, this is exactly what can be observed (see Fig. 6.8). The transition density matrix is localised on both ends of the rod. However, as soon as a truncation to the response density matrix is introduced, the lowest excitation of the system collapses to a state where the transition density is fully localised on the As terminated end of the rod and the excitation energy is pushed up by 0.33 eV.

The dependence of the energy of the localised excited state on the response density matrix truncation radius is plotted in Fig. 6.9. Choosing a density matrix truncation radius of $40\,a_0$ introduces an error of less than 5 meV compared to the converged value for this excited state, suggesting that it is indeed possible to represent localised excited states with a truncated density matrix and thus to compute them in linear-scaling effort.

The remarkable feature of the truncation of the density matrix in this scenario is that it is not introduced to simply improve the scaling as in ground-state DFT. The choice of how to truncate the response density matrix determines what kind of excitations can be represented. This becomes very apparent in the case of the polar nanorod and can be

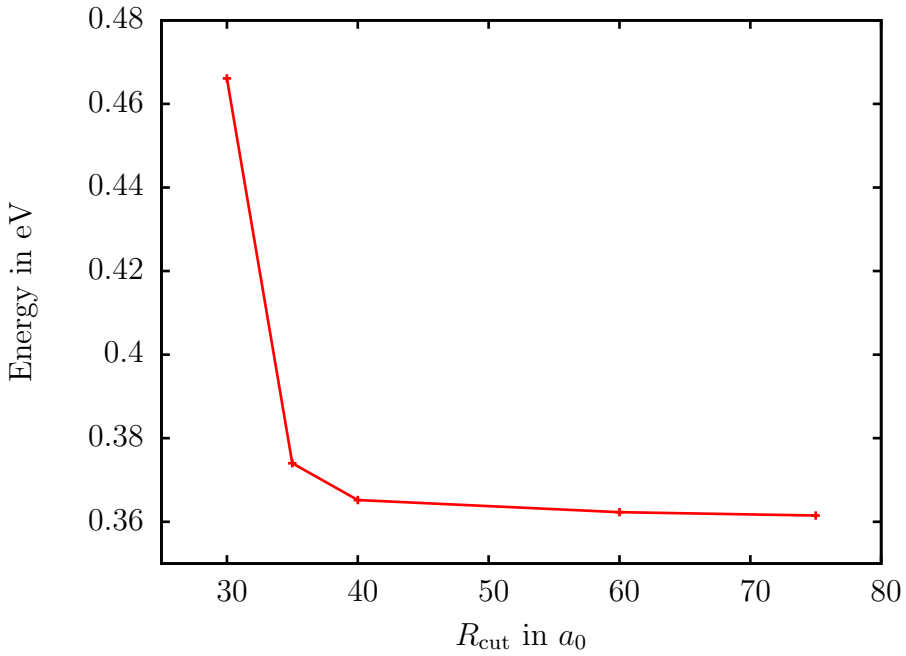


Figure 6.9: Lowest excitation energy of a GaAs nanorod as converged with different response density matrix trunctations (figure adapted from [99]).

understood by some more careful analysis of the structure of the response density matrix. For this purpose the atoms in some region A at LUMO end of the nanorod are labelled by the subscript A_i and the atoms localised at some region B on the HOMO end of the nanorod as B_j. Let furthermore $\{\chi_\alpha^{A_i}\}$ denote the set of all conduction NGWFs centered on atom A_i. Then for the charge transfer state, the response density matrix must be expressible as

$$\rho^{\{1\}}(\mathbf{r},\mathbf{r}') = \sum_{A_i,B_j}\left[\chi_\alpha^{A_i}(\mathbf{r})P^{\{1\}\alpha\beta}_{\{A_i,B_j\}}\phi_\beta^{B_j}(\mathbf{r}')\right] \tag{6.50}$$

Applying the radial cutoff $\rho^{\{1\}}(\mathbf{r},\mathbf{r}') = 0$ for $|\mathbf{r}-\mathbf{r}'| > R_{\text{cut}}$ to the above expression means that the matrix elements $P^{\{1\}\alpha\beta}_{\{A_i,B_j\}} = 0$ for all NGWFs centered on atom A_i and B_j if $|\mathbf{R}_{A_i} - \mathbf{R}_{B_j}| > R_{\text{cut}}$. Therefore, if the regions A and B are separated by more than the chosen cutoff R_{cut}, all matrix elements that are important for representing the charge transfer state between B and A are set to zero and that excitation is effectively removed from the subspace of allowed solutions to the TDDFT algorithm.

Thus, depending on how the sparsity of the response density matrix is chosen, excitations can be converged directly that are not the lowest excitations of the original system. In the case studied here, a truncation of the response density matrix removes all charge transfer excitations along the rod from the subspace of allowed solutions. Since these excitations are precisely the type of excitations that are spuriously low in energy for any local exchange correlation functional, the excitation obtained with the truncated density ma-

trix corresponds to the lowest physical excitation of the system. Thus, the density matrix truncation can be used to effectively remove spurious low energy charge transfer excitations from the system. The remaining excitations are of much more local character and are therefore expected to be treated reasonably accurately by local exchange-correlation functionals. It can be concluded that in the case of the GaAs nanorod treated here, density matrix truncations based on a spherical cutoff can be used to exclude long range excitations from the calculation, while allowing for accurately converging more localised excitations in linear-scaling effort.

It should be noted that the above approach is so effective in this example system because the charge transfer states along the rod do not contribute to the lowest physical excitation localised at the end of the rod. In more complicated systems, one could imagine a situation where a relatively localised state has a small contribution from a state with charge-transfer character. Since this contribution cannot be represented by a truncated density matrix, one would expect to find a discrepancy between the energy of the excited state of the truncated system and the corresponding excited state in the full system. Since there is no way of knowing a priori whether the system at hand possesses a subset of localised excitations well-represented by a truncated response density matrix, the truncation approach can be seen as a semi-empirical way of curing the deficiencies of (semi)-local exchange correlation kernels. However, it should be noted that a failure of the truncation approach can in principle be identified by performing a convergence study like the one shown in Fig. 6.8. If the excitation energy of the truncated system shows sudden drops when increasing the truncation radius, rather than a smooth convergence behaviour, this can be seen as a direct indicator that the excitation has some delocalised charge-transfer contribution in the extended system and thus is not well-represented by a truncated density matrix. The usefulness of a truncation to the response density matrix will be further demonstrated in chapters 8 and 9.

6.9 Representation of the unoccupied space revisited: *In situ* optimisation of a response NGWF basis

In the previous sections, it was demonstrated that the TDDFT method achieves very good results when using a projection onto the entire unoccupied subspace, even if $\{\chi_\alpha\}$ is only optimised for a low energy subset of the conduction space. No matter what choice of projector introduced in 6.3 is used, they all have a feature in common: since the mixing in of unoptimised unoccupied states is allowed, the TDDFT functional $\Omega\left[\{\mathbf{P}_i^{\{1\}}\}\right]$ is no longer strictly minimised with respect to the conduction NGWF expansion coefficients. In section 6.5.1 it became apparent that this factor introduces a starting point dependence

into the algorithm such that slightly different excitation energies are obtained if $\{\chi_\alpha\}$ is optimised for a different number of conduction states. Therefore a very desirable feature of the ground state implementation of ONETEP, the strict minimisation of the energy functional with respect to the NGWF representation, has been lost in the TDDFT implementation. While for the examples considered in this work the low energy excitations are very well represented by the chosen NGWF representation, even when compared to results obtained with rather large Gaussian basis sets, it is in principle desirable to have an equivalent of the two step minimisation procedure of ground state calculations for the TDDFT calculations such that $\{\chi_\alpha\}$ corresponds unambiguously to the ideal representation for a set of N_ω low energy excitations subject to the localisation constraint.

For this purpose, $(\mathbf{S}^\chi)^{-1}$ is used as the projector onto the entire space representable by $\{\chi_\alpha\}$ and the projected conduction Hamiltonian $\mathbf{H}^\chi_{\text{proj}}$ is used in order to force valence-valence style transitions to be unfavourable in energy. Since all reference to any explicit representation of a fixed part of the low energy conduction manifold is now removed from the TDDFT calculation, the conduction optimisation step can be skipped entirely and Ω can be treated as an explicit functional of both $\{\mathbf{P}_i^{\{1\}}\}$ and $\{\chi_\alpha\}$ while keeping $\{\phi_\alpha\}$ and $\mathbf{P}^{\{v\}}$ fixed. In this scenario, it is more appropriate to think of $\{\chi_\alpha\}$ as generalised response NGWFs rather than conduction NGWFs, because they are no longer optimised to represent a certain energy window of the conduction space but rather to ideally represent the response density matrices $\{\mathbf{P}_i^{\{1\}}\}$.

One can then follow a very similar two step minimisation procedure to the one employed in ground-state DFT. Starting with some initial guess for $\{\chi_\alpha\}$, the projector $(\mathbf{S}^\chi)^{-1}$ and the Hamiltonian $\mathbf{H}^\chi_{\text{proj}}$ are built in the given representation. Keeping $\{\chi_\alpha\}$ fixed, Ω is then minimised for a set of $\{\mathbf{P}_i^{\{1\}}\}$ in exactly the same way as described earlier in this chapter. At the end of the conjugate gradient optimisation, the TDDFT energy for the set of $\{\mathbf{P}_i^{\{1\}}\}$ obeying the appropriate the normalisation and invariance constraint can be written in a compact form

$$\begin{aligned}\Omega\left[\{\mathbf{P}_i^{\{1\}}\},\{\chi_\alpha\}\right] &= \sum_i \frac{\text{Tr}\left[\mathbf{H}^\chi_{\text{proj}}\mathbf{P}_i^{\{\text{elec}\}}\right] - \text{Tr}\left[\mathbf{S}^\chi\mathbf{P}_i^{\{1\}}\mathbf{H}^\chi\mathbf{P}_i^{\{1\}\dagger}\right]}{\text{Tr}\left[\mathbf{S}^\chi\mathbf{P}_i^{\{\text{elec}\}}\right]} \\ &+ \sum_i \frac{\text{Tr}\left[\mathbf{P}_i^{\{1\}\dagger}\mathbf{V}^{\{1\}\chi\phi}_{\text{SCF}}\left[\mathbf{P}_i^{\{1\}}\right]\right]}{\text{Tr}\left[\mathbf{S}^\chi\mathbf{P}_i^{\{\text{elec}\}}\right]}\end{aligned} \tag{6.51}$$

One can then straightforwardly calculate the gradient of $\Omega\left[\{\mathbf{P}_i^{\{1\}}\},\{\chi_\alpha\}\right]$ with respect

to χ_β, yielding

$$\begin{aligned}\frac{1}{2}\left(\frac{\partial\Omega}{\partial\chi_\beta}\right)_j &= \hat{H}|\chi_\alpha\rangle P_j^{\{\mathrm{elec}\}\alpha\beta} - |\phi_\alpha\rangle\left(P^{\{v\}}H^\phi P^{\{v\}}S^{\phi\chi}P_j^{\{\mathrm{elec}\}}\right)^{\alpha\beta} \\ &+ \lambda|\phi_\alpha\rangle\left(P^{\{v\}}S^\phi P^{\{v\}}S^{\phi\chi}P_j^{\{\mathrm{elec}\}}\right)^{\alpha\beta} - |\chi_\alpha\rangle\left(P_j^{\{1\}}H^\phi P_j^{\{1\}\dagger}\right)^{\alpha\beta} \\ &+ \hat{V}_{\mathrm{SCF}}^{\{1\}}\left[\mathbf{P}_j^{\{1\}}\right]|\phi_\alpha\rangle P^{\{1\}\beta\alpha} - \mathrm{Tr}\left[\mathbf{P}_i^{\{1\}\dagger}\mathbf{S}^\chi\mathbf{q}_j^{\chi\phi}\mathbf{S}^\phi\right]|\chi_\alpha\rangle P_j^{\{\mathrm{elec}\}\alpha\beta}\end{aligned}$$

Here, the fact is used that the response density matrices are normalised.

Then the contravariant gradient of the TDDFT energy functional with respect to the psinc expansion coefficients $C_{m\alpha}$ of the NGWFs $\{\chi_\beta\}$ can be expressed via

$$g_m^\beta = w\sum_j\left[\left(\frac{\partial\Omega}{\partial\chi_\beta}\right)_j\right]_{\mathbf{r}=\mathbf{r}_m} \tag{6.52}$$

where w is the grid point weighting and $\mathbf{r}_m$ is the grid point associated with psinc function m. The contravariant gradient can be transformed into a covariant by multiplying it with $\mathbf{S}^\chi$, which can then be used to update $\{\chi_\alpha\}$. Thus, a two step minimisation similar to the ground state version can be carried out without any need to perform the an explicit conduction NGWF optimisation before starting the TDDFT calculation. When $\Omega\left[\{\mathbf{P}_i^{\{1\}}\},\{\chi_\alpha\}\right]$ is minimised both with respect to the response NGWFs $\{\chi_\alpha\}$ and the response density matrices $\{\mathbf{P}_i^{\{1\}}\}$, an optimal representation of the TDDFT response for the N_ω lowest excitations has been found.

While the above outline details in principle how to perform the minimisation of the TDDFT energy with respect to the generalised response NGWFs $\{\chi_\alpha\}$, carrying out the minimisation in an efficient way is not trivial and will be left for future work. For the remainder of this work, the conduction NGWFs are taken to be fixed and the joint set and the idempotent projector onto the entire unoccupied subspace are used, as they are found to yield the most consistent results of all the fixed NGWF set approaches outlined in this section.

Chapter 7

Linear-scaling TDDFT within the PAW formalism

In the previous chapter, the linear-scaling TDDFT method was introduced for the case of norm-conserving pseudopotentials. In this chapter the focus is on expanding the algorithm to make it compatible with the linear-scaling projector augmented-wave (PAW) formalism available in ground state and conduction calculations with `ONETEP` (see section 5.2.2). The correct workings of the formalism are then demonstrated on a number of small test molecules, as well as the physically motivated system of a Cadmium Sulphide (CdS) nanocrystal under pressure.

7.1 Derivation of the formalism

In analogy to the conventional implementation of PAW in `ONETEP`, one would like to write the all-electron equivalent of the TDDFT excitation energy as a functional of the soft pseudo response density matrix $\mathbf{P}^{\{1\}}$ that is well represented by the conduction and valence NGWFs. The total all-electron response density matrix can be written in operator form in terms of the pseudo density matrix, the PAW projectors $|\tilde{p}^i\rangle$ and the all-electron and pseudo partial waves $|\varphi_i\rangle$ and $|\tilde{\varphi}_i\rangle$:

$$\rho^{\{1\}} = |\chi_\alpha\rangle P^{\{1\}\alpha\beta}\langle\phi_\beta| + \sum_{ij}\left\{|\varphi_i\rangle\langle\varphi_j| - |\tilde{\varphi}_i\rangle\langle\tilde{\varphi}_j|\right\}\langle\tilde{p}^i|\chi_\alpha\rangle P^{\{1\}\alpha\beta}\langle\phi_\beta|\tilde{p}^j\rangle \tag{7.1}$$

As with the ground state and conduction density matrices, the response density matrix thus splits into a soft pseudo part defined everywhere in the system and an all-electron part defined only in the PAW spheres around the atoms. To express the sphere terms, one needs the projector-NGWF overlap matrices $Q_\alpha^{\{\chi\}i} = \langle\chi_\alpha|\tilde{p}^i\rangle$ and $Q_\beta^{\{\phi\}j} = \langle\phi_\beta|\tilde{p}^j\rangle$. These matrices can be precomputed at the beginning of the TDDFT calculation by using the NGWF sets optimised in the ground state and conduction calculation. Together they can be used to express the projector matrix $\rho^{\{1\}ij}$ with

$$\rho^{\{1\}ij} = \langle\tilde{p}^i|\chi_\alpha\rangle P^{\{1\}\alpha\beta}\langle\phi_\beta|\tilde{p}^j\rangle = \left(Q_\alpha^{\{\chi\}i}\right)^\dagger P^{\{1\}\alpha\beta} Q_\beta^{\{\phi\}j}. \tag{7.2}$$

For the PAW formalism, the all-electron excitation energy E_{TDDFT} is written as a functional of the PS density matrix $\mathbf{P}^{\{1\}}$ and is minimised under the normalisation constraint and invariance constraints discussed in the previous chapter. However, in the PAW formalism, these constraints are placed on the all electron response density matrix and not the pseudised version $\mathbf{P}^{\{1\}}$. It is thus required that the all-electron response density matrix in operator form fulfills:

$$\begin{aligned} \mathrm{Tr}\left[\rho^{\{1\}\dagger}\rho^{\{1\}}\right] &= \mathrm{Tr}\left[\rho^{\{1\}}\rho^{\{1\}\dagger}\right] = 1 & (7.3)\\ \rho^{\{c\}}\rho^{\{1\}}\rho^{\{v\}} &= \rho^{\{1\}} & (7.4) \end{aligned}$$

where $\rho^{\{v\}}$ and $\rho^{\{c\}}$ are the all-electron conduction and valence density matrices in operator form. Just like in conventional PAW, it is easy to show that the above conditions can be written in terms of the pseudo response density matrix only. Redefining the overlap matrices $\mathbf{S}^{\chi}$ and $\mathbf{S}^{\phi}$ in the same fashion as in conventional PAW, with

$$S^{\phi}_{\alpha\beta} = \langle\phi_\alpha| \left(1 + \sum_{ij} |\tilde{p}^i\rangle(\langle\varphi_i|\varphi_j\rangle - \langle\tilde{\varphi}_i|\tilde{\varphi}_j\rangle)\langle\tilde{p}^j|\right) |\phi_\beta\rangle \tag{7.5}$$

$$S^{\chi}_{\alpha\beta} = \langle\chi_\alpha| \left(1 + \sum_{ij} |\tilde{p}^i\rangle(\langle\varphi_i|\varphi_j\rangle - \langle\tilde{\varphi}_i|\tilde{\varphi}_j\rangle)\langle\tilde{p}^j|\right) |\chi_\beta\rangle \tag{7.6}$$

the normalisation and invariance constraint on the all-electron response density matrix can be conveniently written as

$$\begin{aligned} \mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\right] &= 1 & (7.7)\\ \mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}} &= \mathbf{P}^{\{1\}}. & (7.8) \end{aligned}$$

Note that the above constraints have exactly the same form as in the norm-conserving pseudopotential case, and all references to all-electron terms are wrapped up in the redefinition of the overlap matrices.

In the norm-conserving approach, the TDDFT excitation energy is a functional of the the response density $\rho^{\{1\}}(\mathbf{r})$. The excitation energy can be split into three different parts, with

$$E_{\mathrm{TDDFT}}\left[\rho^{\{1\}}\right] = E_{\mathrm{diag}}\left[\rho^{\{1\}}\right] + 2E_{\mathrm{H}}\left[\rho^{\{1\}}\right] + 2E_{\mathrm{xc}}\left[\rho^{\{1\}}\right] \tag{7.9}$$

where

$$E_{\text{diag}}\left[\rho^{\{1\}}\right] = \text{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{H}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\right] - \text{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{H}^{\phi}\right] \tag{7.10}$$

$$E_{\text{H}}\left[\rho^{\{1\}}\right] = \int\int \text{d}^3\,\text{d}^3r'\,\frac{\rho^{\{1\}}(\mathbf{r})\rho^{\{1\}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{7.11}$$

$$E_{\text{xc}}\left[\rho^{\{1\}}\right] = \int \text{d}^3r\,\text{d}^3r'\,\delta(\mathbf{r}-\mathbf{r}')\rho^{\{1\}}(\mathbf{r})f_{\text{xc}}(\mathbf{r})\rho^{\{1\}}(\mathbf{r}'). \tag{7.12}$$

Here, f_{xc} is assumed to be (semi)-local and any explicit spin dependence is again dropped for convenience. In the following sections, each of the three terms will be considered individually and the changes required in writing them as all-electron equivalents that are functionals of the soft pseudo response density matrix are addressed.

7.1.1 The diagonal term

The conceptually most simple contribution to the TDDFT excitation energy is the diagonal term of Kohn–Sham eigenvalue differences. First, it becomes convenient to rewrite $E_{\text{diag}}\left[\rho^{\{1\}}\right]$ in terms of the effective electron and hole density matrix by making use of the invariance of the trace under cyclic permutations:

$$\begin{aligned} E_{\text{diag}}\left[\rho^{\{1\}}\right] &= \text{Tr}\left[\mathbf{H}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{1\}\dagger}\right] - \text{Tr}\left[\mathbf{H}^{\phi}\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\right] \\ &= \text{Tr}\left[\mathbf{H}^{\chi}\mathbf{P}^{\{\text{elec}\}}\right] - \text{Tr}\left[\mathbf{H}^{\phi}\mathbf{P}^{\{\text{hole}\}}\right]. \end{aligned} \tag{7.13}$$

In other words, the diagonal term in the TDDFT algorithm is due to the difference between the Kohn–Sham bandstructure energy of the electron and the hole. In the case of a PAW treatment of the TDDFT energy, the Kohn–Sham bandstructure energies of the electron and the hole need to be replaced by their all-electron equivalents and need to be expressed in terms of the soft pseudo versions of the electron- and the hole density matrices. However, the required all-electron version of the Kohn–Sham bandstructure energy for an arbitrary density matrix is exactly what is required in the ground state and conduction state optimisation using the PAW method. Thus, the diagonal part of the TDDFT energy is obtained by simply replacing the conduction and valence Hamiltonians by their PAW equivalents such that

$$H^{\phi}_{\alpha\beta} = \langle\phi_\alpha|\left(-\frac{1}{2}\nabla^2 + \tilde{V}_{\text{KS}} + \sum_{ij}|\tilde{p}^i\rangle D_{ij}\langle\tilde{p}^j|\right)|\phi_\beta\rangle \tag{7.14}$$

$$H^{\chi}_{\alpha\beta} = \langle\chi_\alpha|\left(-\frac{1}{2}\nabla^2 + \tilde{V}_{\text{KS}} + \sum_{ij}|\tilde{p}^i\rangle D_{ij}\langle\tilde{p}^j|\right)|\chi_\beta\rangle \tag{7.15}$$

with D_{ij} being the non-local PAW correction term to the Hamiltonian (see section 5.2.2). The gradient of $E_{\text{diag}}\left[\rho^{\{1\}}\right]$ with respect to $\mathbf{P}^{\{1\}}$ can then be found trivially, since $\mathbf{H}^{\chi}$ and $\mathbf{H}^{\phi}$ are not functionals of $\mathbf{P}^{\{1\}}$. Thus it is found that

$$\frac{\mathrm{d}E_{\text{diag}}}{\mathrm{d}\mathbf{P}^{\{1\}}} = \frac{\partial E_{\text{diag}}}{\partial \mathbf{P}^{\{1\}}} = 2\left(\mathbf{H}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi} - \mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{H}^{\phi}\right) \tag{7.16}$$

and that the gradient of the diagonal part of the TDDFT energy can be evaluated trivially after constructing the PAW-corrected conduction and valence Hamiltonians.

7.1.2 The Hartree term

To calculate the all-electron equivalent of the Hartree term due to response density $\rho^{\{1\}}(\mathbf{r})$, a similar partitioning technique to the one used in conventional ground-state PAW is considered. The AE response density can be written as

$$\rho^{\{1\}} = \left(\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}\right) + \left(\rho^{\{1\}}\right)^{(1)} - \left(\left(\tilde{\rho}^{\{1\}}\right)^{(1)} + \hat{\rho}^{\{1\}}\right) \tag{7.17}$$

where $\tilde{\rho}^{\{1\}}$ is the soft part of the response density, $\left(\rho^{\{1\}}\right)^{(1)}$ is the hard all-electron density defined in the sphere regions and $\left(\tilde{\rho}^{\{1\}}\right)^{(1)}$ is the density in the sphere regions defined via the pseudo partial waves. The densities within the augmentation sphere regions are constructed from the projector density matrix $\rho^{\{1\}ij}$ in exactly the same way as in the ground state PAW implementation. $\hat{\rho}^{\{1\}}$ again denotes the compensation density chosen in such a way that the Coulomb potential in one sphere has no contributions in the other sphere. As in the ground state version of PAW, adding the compensation density is crucial to preserve locality of the PAW sphere terms. In close analogy to the ground state calculation, $\hat{\rho}^{\{1\}}$ is simply given by

$$\hat{\rho}^{\{1\}}(\mathbf{r}) = \sum_{ijLM} \rho^{\{1\}ij}\hat{Q}_{ij}^{LM}(\mathbf{r}) \tag{7.18}$$

where $\hat{Q}_{ij}^{LM}(\mathbf{r})$ follows the same form as in the ground state calculation and can be trivially preconstructed using the AE and PS partial waves. Thus, in close analogy to the ground state calculations, the PAW Hartree energy of the response density can be split into three parts: the soft PS part defined everywhere and a AE and a PS part only defined within the PAW spheres. The Hartree energy is then given by

$$E_{\text{H}}\left[\rho^{\{1\}}\right] = E_{\text{H}}\left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}\right] + E_{\text{H}}\left[\left(\rho^{\{1\}}\right)^{(1)}\right] - E_{\text{H}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)} + \hat{\rho}^{\{1\}}\right]. \tag{7.19}$$

The gradient of the Hartree energy with respect to the PS response density matrix $\mathbf{P}^{\{1\}}$ can then be found via

$$\frac{\mathrm{d}E_{\mathrm{H}}}{\mathrm{d}\mathbf{P}^{\{1\}}} = \frac{\partial E_{\mathrm{H}}}{\partial \mathbf{P}^{\{1\}}} + \sum_{ij} \frac{\partial E_{\mathrm{H}}}{\partial \rho^{\{1\}ij}} \frac{\partial \rho^{\{1\}ij}}{\partial \mathbf{P}^{\{1\}}}. \tag{7.20}$$

Following similar steps as in the ground state case, the gradient of the Hartree TDDFT energy can be written

$$\begin{aligned}
\frac{1}{2}\left(\frac{\mathrm{d}E_{\mathrm{H}}}{\mathrm{d}\mathbf{P}^{\{1\}}}\right)_{\alpha\beta} &= \langle\chi_\alpha|V_{\mathrm{H}}\left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}\right]|\phi_\beta\rangle \\
&+ \sum_{ij}\langle\chi_\alpha|\tilde{p}^i\rangle \sum_{LM}\int \mathrm{d}^3r\, \hat{Q}_{ij}^{LM}(\mathbf{r}) V_{\mathrm{H}}\left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}\right](\mathbf{r})\langle\tilde{p}^j|\phi_\beta\rangle \\
&+ \sum_{ij}\langle\chi_\alpha|\tilde{p}^i\rangle\langle\varphi_i|V_{\mathrm{H}}\left[\left(\rho^{\{1\}}\right)^{(1)}\right]|\varphi_j\rangle\langle\tilde{p}^j|\phi_\beta\rangle \\
&- \sum_{ij}\langle\chi_\alpha|\tilde{p}^i\rangle \sum_{LM}\int_\Omega \mathrm{d}^3r\, \hat{Q}_{ij}^{LM}(\mathbf{r}) V_{\mathrm{H}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)} + \hat{\rho}^{\{1\}}\right](\mathbf{r})\langle\tilde{p}^j|\phi_\beta\rangle \\
&- \sum_{ij}\langle\chi_\alpha|\tilde{p}^i\rangle\langle\tilde{\varphi}_i|V_{\mathrm{H}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)} + \hat{\rho}^{\{1\}}\right]|\tilde{\varphi}_j\rangle\langle\tilde{p}^j|\phi_\beta\rangle
\end{aligned} \tag{7.21}$$

where integral that is only performed in PAW sphere are again denoted with the subscript Ω [35].

In the spirit of ground state PAW, this equation can be written in a compact form such that

$$\left(\frac{\mathrm{d}E_{\mathrm{H}}}{\mathrm{d}\mathbf{P}^{\{1\}}}\right)_{\alpha\beta} = 2\left(\tilde{V}_{\mathrm{H}}^{\chi\phi}\left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}\right]\right)_{\alpha\beta} + 2\sum_{ij} Q_\alpha^{\{\chi\}i} D_{ij}^{\{\mathrm{H}\}} \left(Q_\beta^{\{\phi\}j}\right)^\dagger \tag{7.22}$$

with the non-local correction term $D_{ij}^{\{\mathrm{H}\}}$ given by

$$\begin{aligned}
D_{ij}^{\{\mathrm{H}\}} &= \sum_{LM}\int \mathrm{d}^3r\, \hat{Q}_{ij}^{LM}(\mathbf{r}) V_{\mathrm{H}}\left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}\right](\mathbf{r}) + \langle\varphi_i|V_{\mathrm{H}}\left[\left(\rho^{\{1\}}\right)^{(1)}\right]|\varphi_j\rangle \\
&- \sum_{LM}\int_\Omega \mathrm{d}^3r\, \hat{Q}_{ij}^{LM}(\mathbf{r}) V_{\mathrm{H}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)} + \hat{\rho}^{\{1\}}\right](\mathbf{r}) - \langle\tilde{\varphi}_i|V_{\mathrm{H}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{1} + \hat{\rho}^{\{1\}}\right]|\tilde{\varphi}_j\rangle
\end{aligned} \tag{7.23}$$

Note that this decomposition is exactly equivalent to the way the ground state Hartree potential was decomposed in conventional PAW. This means that the last three terms of the above expression can be evaluated in the same way as in the ground state case so that

[35] Again, as in the ground state PAW case, the integral involving the soft PS response density and $\hat{Q}_{ij}^{LM}(\mathbf{r})$ is formally evaluated on the entire real space grid. However, in practice, it can be calculated in a small augmentation box around each atom due to the localisation of $\hat{Q}_{ij}^{LM}(\mathbf{r})$.

the full expression reduces to

$$D_{ij}^{\{\mathrm{H}\}} = \sum_{LM} \int \mathrm{d}^3 r \, \hat{Q}_{ij}^{LM}(\mathbf{r}) V_{\mathrm{H}} \left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}} \right] + \sum_{kl} e_{ijkl} \rho^{\{1\}kl} \tag{7.24}$$

where e_{ijkl} is the same rank 4 tensor that gets precomputed in the ground state and conduction state calculations when reading the PAW datasets from file.

Thus the gradient of the Hartree energy for a given response density $\rho^{\{1\}}(\mathbf{r})$ can be computed in a number of straightforward steps analogous to the ground state evaluation of the Hartree energy. First, the compensation density $\hat{\rho}^{\{1\}}$ is constructed from the response density matrix and the PAW projectors. Then the soft PS part of the Hartree term is constructed as usual, but one now adds the compensation density to the soft response density $\tilde{\rho}^{\{1\}}$. The PAW correction terms to the soft Hartree potential are then simply constructed via equation (7.24) and added to the PS part.

7.1.3 The exchange-correlation kernel

For the exchange-correlation part, one can again divide the energy gradient into a PS part evaluated on the normal simulation cell grid and a PAW correction term. Note however, that in this case, both the response density $\rho^{\{1\}}$ and the ground state density ρ need to be partitioned in the usual PAW fashion, as the exchange-correlation kernel f_{xc} needs to be evaluated at the ground state density. Furthermore, in order to simplify a number of expressions, it is helpful to reintroduce the finite-difference approximation to the exchange-correlation term such that the soft PS part of the exchange-correlation potential can be written as

$$\begin{aligned} & \int \mathrm{d}^3 r' \delta(\mathbf{r} - \mathbf{r}') f_{\mathrm{xc}} \left[\tilde{\rho} + \hat{\rho} + \tilde{\rho}_c \right] (\mathbf{r}') \left(\tilde{\rho}^{\{1\}}(\mathbf{r}') + \hat{\rho}^{\{1\}}(\mathbf{r}') \right) \\ \approx & \frac{V_{\mathrm{xc}} \left[\tilde{\rho} + \hat{\rho} + \tilde{\rho}_c + \epsilon(\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}) \right] (\mathbf{r}) - V_{\mathrm{xc}} \left[\tilde{\rho} + \hat{\rho} + \tilde{\rho}_c - \epsilon(\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}) \right] (\mathbf{r})}{2\epsilon} \\ = & \Delta V_{\mathrm{xc}} \left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}} \right] (\mathbf{r}) \end{aligned} \tag{7.25}$$

Similarly, the soft and the AE parts in the PAW spheres can be evaluated from

$$\Delta V_{\mathrm{xc}}\left[\left(\rho^{\{1\}}\right)^{(1)}\right](\mathbf{r}) = \frac{V_{\mathrm{xc}}\left[\rho^{(1)}+\rho_c+\epsilon\left(\rho^{\{1\}}\right)^{(1)}\right](\mathbf{r})}{2\epsilon} - \frac{V_{\mathrm{xc}}\left[\rho^{(1)}+\rho_c-\epsilon\left(\rho^{\{1\}}\right)^{(1)}\right](\mathbf{r})}{2\epsilon} \tag{7.26}$$

$$\Delta V_{\mathrm{xc}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)}+\hat{\rho}^{\{1\}}\right](\mathbf{r}) = \frac{V_{\mathrm{xc}}\left[\left(\tilde{\rho}^{\{v\}}\right)^{(1)}+\hat{\rho}^{\{v\}}+\tilde{\rho}_c+\epsilon\left(\tilde{\rho}^{\{1\}}+\hat{\rho}^{\{1\}}\right)^{(1)}\right](\mathbf{r})}{2\epsilon} - \frac{V_{\mathrm{xc}}\left[\left(\tilde{\rho}^{\{v\}}\right)^{(1)}+\hat{\rho}^{\{v\}}+\tilde{\rho}_c-\epsilon\left(\tilde{\rho}^{\{1\}}+\hat{\rho}^{\{1\}}\right)^{(1)}\right](\mathbf{r})}{2\epsilon} \tag{7.27}$$

Using the finite difference approximation above facilitates the calculation of the PAW sphere correction terms as they can all be written in terms of the exchange-correlation potential only and thus become very similar to the terms that need to be evaluated in ground state PAW calculations. Previous PAW implementations in a plane-wave TDDFT code have made use of the same finite difference approximation and found the scheme to be very insensitive to the finite-difference parameter ϵ used [138][36].

Following similar steps to the way the gradient of the Hartree energy was constructed in the previous section and using the finite difference approximation for the TDDFT kernel it is found that

$$\left(\frac{\mathrm{d}E_{\mathrm{xc}}}{\mathrm{d}\mathbf{P}^{\{1\}}}\right)_{\alpha\beta} = 2\langle\chi_\alpha|\Delta V_{\mathrm{xc}}\left[\tilde{\rho}^{\{1\}}+\hat{\rho}^{\{1\}}\right]|\phi_\beta\rangle + 2\sum_{ij} Q_\alpha^{\{\chi\}i} D_{ij}^{\{\mathrm{xc}\}} \left(Q_\beta^{\{\phi\}j}\right)^\dagger \tag{7.28}$$

where

$$\begin{aligned} D_{ij}^{\{\mathrm{xc}\}} &= \sum_{LM}\int \mathrm{d}^3r\, \hat{Q}_{ij}^{LM}(\mathbf{r})\Delta V_{\mathrm{xc}}\left[\tilde{\rho}^{\{1\}}+\hat{\rho}^{\{1\}}\right](\mathbf{r}) \\ &+ \langle\varphi_i|\Delta V_{\mathrm{xc}}\left[\left(\rho^{\{1\}}\right)^{(1)}\right]|\varphi_j\rangle - \langle\tilde{\varphi}_i|\Delta V_{\mathrm{xc}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)}+\hat{\rho}^{\{1\}}\right]|\tilde{\varphi}_j\rangle \\ &- \sum_{LM}\int_\Omega \mathrm{d}^3r\, \Delta V_{\mathrm{xc}}\left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)}+\hat{\rho}^{\{1\}}\right](\mathbf{r})\hat{Q}_{ij}^{LM}(\mathbf{r}) \end{aligned} \tag{7.29}$$

The above expression is very similar to the PAW correction obtained in the ground state case, with the difference that any V_{xc} term is replaced by the corresponding finite-difference term ΔV_{xc}. Just like in the ground state case, the sphere terms can be evaluated

[36] In ONETEP, if ALDA functionals are used, the soft part of the exchange-correlation term in TDDFT is actually evaluated using the analytic expression for f_{xc}, while the sphere terms are treated in a finite difference approximation. The errors introduced through this slight inconsistency in the way different terms are treated are found to be negligible.

by expressing ΔV_{xc} in terms of its moments $\Delta V_{LM}^{\mathrm{xc}}$ such that

$$\Delta V_{\mathrm{xc}}[\rho](\mathbf{r}) = \sum_{LM} \Delta V_{LM}^{\mathrm{xc}}[\rho](r) S_{LM}(\hat{\mathbf{r}}). \tag{7.30}$$

Again, as in the ground state case, these moments are constructed by approximating the appropriate density as a second order Taylor expansion around its spherically symmetric moment with $L, M = 0$.

Once the moments are calculated for the three different ΔV_{xc} terms, the PAW correction term can be written as:

$$\begin{aligned} D_{ij}^{\{\mathrm{xc}\}} &= \sum_{LM} \int \mathrm{d}^3 r\, \hat{Q}_{ij}^{LM}(\mathbf{r}) \Delta V_{\mathrm{xc}} \left[\tilde{\rho}^{\{1\}} + \hat{\rho}^{\{1\}}\right](r) \\ &+ \sum_{LM} \mathcal{G}_{l_i m_i l_j m_j}^{LM} \int_0^{r_c} \Delta V_{LM}^{\mathrm{xc}} \left[\left(\rho^{\{1\}}\right)^{(1)}\right](r) \varphi_{n_i l_i}(r) \varphi_{n_j l_j}(r) \mathrm{d}r \\ &- \sum_{LM} \mathcal{G}_{l_i m_i l_j m_j}^{LM} \int_0^{r_c} \Delta V_{LM}^{\mathrm{xc}} \left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)} + \hat{\rho}^{\{1\}}\right](r) \tilde{\varphi}_{n_i l_i}(r) \tilde{\varphi}_{n_j l_j}(r) \mathrm{d}r \\ &- \sum_{LM} \mathcal{G}_{l_i m_i l_j m_j}^{LM} \int_0^{r_c} \Delta V_{LM}^{\mathrm{xc}} \left[\left(\tilde{\rho}^{\{1\}}\right)^{(1)} + \hat{\rho}^{\{1\}}\right](r) n_{n_i l_i n_j l_j}^{L} g_L(r) r^2 \mathrm{d}r \end{aligned} \tag{7.31}$$

where the above integrations can be conveniently performed on the radial grid in the PAW spheres around the atoms. Here, $\mathcal{G}_{l_i m_i l_j m_j}^{LM}$ again denotes the set of precomputed Gaunt coefficients and the definition of the shape function $g_L(r)$ and the coefficients $n_{n_i l_i n_j l_j}^{L}$ are unchanged from the ground-state PAW case (see 5.2.2).

Since only spin-degenerate ground states are considered in this work, reintroducing the spin dependence into the exchange correlation kernel is relatively trivial and can be performed in the same way as in the conventional TDDFT implementation. One again obtains two different exchange correlation terms, depending on whether singlets or triplets are calculated, with

$$\Delta V_{\mathrm{xc}}^{\mathrm{triplet}}[\rho^{\{1\}}] = \frac{1}{2}\left(\Delta V_{\mathrm{xc}}[\rho^{\{1\}}] - \Delta V_{\mathrm{c}}[\rho^{\{1\}}]\right) \tag{7.32}$$

$$\Delta V_{\mathrm{xc}}^{\mathrm{singlet}}[\rho^{\{1\}}] = \frac{1}{2}\left(\Delta V_{\mathrm{xc}}[\rho^{\{1\}}] + \Delta V_{\mathrm{c}}[\rho^{\{1\}}]\right) \tag{7.33}$$

where the exchange-correlation term ΔV_{xc} and the correlation-only term ΔV_{c} are then spin-degenerate quantities. Note that in the case of the triplet excitations, the Hartree term of the previous section is set to zero in agreement with the conventional TDDFT algorithm in the norm-conserving approach.

7.1.4 The algorithm

Taking the results from the previous sections the gradient of the TDDFT energy with respect to the PS density kernel $\mathbf{P}^{\{1\}}$ can be written as[37]

$$\begin{aligned}\left(\frac{\mathrm{d}E_{\mathrm{TDDFT}}}{\mathbf{P}^{\{1\}}}\right)_{\alpha\beta} &= 2\left(H^{\chi}P^{\{1\}}S^{\phi} - S^{\chi}P^{\{1\}}H^{\phi}\right)_{\alpha\beta} \\ &+ 2\left(\tilde{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}\left[\tilde{\rho}^{\{1\}}+\hat{\rho}^{\{1\}}\right]\right)_{\alpha\beta} \\ &+ 4\sum_{ij} Q_{\alpha}^{\{\chi\}i}\left(D_{ij}^{\{\mathrm{H}\}}+D_{ij}^{\{\mathrm{xc}\}}\right)\left(Q_{\beta}^{\{\phi\}j}\right)^{\dagger} \end{aligned} \tag{7.34}$$

A contravariant gradient is required that automatically preserves the invariance constraint on the response density matrix to form a PAW equivalent of the quantity $\mathbf{q}^{\chi\phi}$ in the norm-conserving approach. This can be achieved by taking equation (7.34), mutliplying by the PS conduction density matrix from the left and the valence density matrix from the right and dividing by a factor of 2. After using the idempotency constraint on the conduction and valence density matrix, one arrives at

$$\begin{aligned}\left(q_{\mathrm{PAW}}^{\chi\phi}\right)^{\alpha\beta} &= \left(P^{\{c\}}H^{\chi}P^{\{1\}} - P^{\{1\}}H^{\phi}P^{\{v\}}\right)^{\alpha\beta} \\ &+ \left(P^{\{c\}}\left\{\tilde{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}\left[\tilde{\rho}^{\{1\}}+\hat{\rho}^{\{1\}}\right] + 2\sum_{ij} Q^{\{\chi\}i}\left(D_{ij}^{\{\mathrm{H}\}}+D_{ij}^{\{\mathrm{xc}\}}\right)\left(Q^{\{\phi\}j}\right)^{\dagger}\right\}P^{\{v\}}\right)^{\alpha\beta}\end{aligned} \tag{7.35}$$

and the PAW TDDFT energy can then be written in the compact form

$$E_{\mathrm{TDDFT}}\left[\rho^{\{1\}}\right] = \omega = \frac{\mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{q}_{\mathrm{PAW}}^{\alpha\beta}\mathbf{S}^{\phi}\right]}{\mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\right]}. \tag{7.36}$$

Note that the PAW-corrected TDDFT operator $\mathbf{q}_{\mathrm{PAW}}^{\chi\phi}$ has a very similar form to the conventional $\mathbf{q}^{\chi\phi}$ in the norm-conserving approach. The only difference is that the Hamiltonians and overlap matrices in the conduction and valence NGWF representations are replaced by their PAW equivalents, while $\tilde{V}_{\mathrm{SCF}}^{\{1\}}(\mathbf{r})$ is evaluated for the soft PS density and the compensation density on the standard simulation cell grid and is then enhanced by the in-sphere PAW correction terms $D_{ij}^{\{\mathrm{H}\}}$ and $D_{ij}^{\{\mathrm{xc}\}}$.

Thus, only minimal changes to the conventional TDDFT algorithm are required for the PAW formalism. While the conduction and valence Hamiltonians $\mathbf{H}^{\chi}$ and $\mathbf{H}^{\phi}$ have to be constructed only once at the beginning of the calculation, each time the TDDFT

[37]Note that a factor of two has been absorbed into $\tilde{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}\left[\tilde{\rho}^{\{1\}}+\hat{\rho}^{\{1\}}\right](\mathbf{r})$ in order to keep its definition fully analogous to the norm-conserving case.

gradient is evaluated a new $\hat{\rho}^{\{1\}}$ and new PAW sphere terms $D_{ij}^{\{\mathrm{H}\}}$ and $D_{ij}^{\{\mathrm{xc}\}}$ have to be constructed. However, note that the routines to calculate $\hat{\rho}^{\{1\}}$, $D_{ij}^{\{\mathrm{H}\}}$ and $D_{ij}^{\{\mathrm{xc}\}}$ can be reused from standard ground state PAW calculations. $D_{ij}^{\{\mathrm{H}\}}$ is evaluated in precisely the same way as in ground state calculations, it is simply computed with a different density kernel $\rho^{\{1\}ij}$, while thanks to the finite-difference technique, the $D_{ij}^{\{\mathrm{xc}\}}$ term can be computed completely analogous to the ground state version, only that now fractions of the response density are added to and subtracted from the appropriate fixed ground state density. Since the sphere terms are fully local in character, they do not alter the computational efficiency of the TDDFT algorithm with system size. Once the PAW-corrected operator $\mathbf{q}_{\mathrm{PAW}}^{\chi\phi}$ is computed for a given guess to the response density matrix, the TDDFT conjugate gradient search direction is evaluated in exactly the same way as in conventional TDDFT calculations.

The only other quantity that needs to be changed for the TDDFT algorithm to be compatible with the PAW formalism is the transition dipole moment that is used to compute the oscillator strength. In the conventional norm-conserving approach it is given by

$$\mu = P^{\{1\}\alpha\beta}\langle\chi_\alpha|\mathbf{r}|\phi_\beta\rangle \tag{7.37}$$

while in the PAW approach it has to be computed for the all-electron response matrix rather than for the soft PS part only. This can be achieved by applying the PAW transformation such that

$$\begin{aligned}\mu &= P^{\{1\}\alpha\beta}\langle\chi_\alpha|\left(\mathbf{r}+\sum_{ij}|\tilde{p}^i\rangle\left\{\langle\varphi_i|\mathbf{r}|\varphi_j\rangle-\langle\tilde{\varphi}_i|\mathbf{r}|\tilde{\varphi}_j\rangle\right\}\langle\tilde{p}^j|\right)|\phi_\beta\rangle &(7.38)\\ &= P^{\{1\}\alpha\beta}\langle\chi_\alpha|\mathbf{r}|\phi_\beta\rangle+\sum_{ij}\rho^{\{1\}ij}\left\{\langle\varphi_i|\mathbf{r}|\varphi_j\rangle-\langle\tilde{\varphi}_i|\mathbf{r}|\tilde{\varphi}_j\rangle\right\} &(7.39)\end{aligned}$$

The PAW correction term can be trivially evaluated inside the PAW spheres and added to the conventional term for the oscillator strength.

7.2 Benchmark tests

A number of benchmark tests are performed to demonstrate that the ONETEP approach can be used to obtain optical spectra for low energy excitations of systems that could not be treated effectively with the norm-conserving pseudopotential approach. For these small test molecules, the benchmark tests to the all-electron results obtained from NWChem [128]. A more realistic test system of a cadmium sulphide (CdS) nanocrystal under pressure is then considered.

Unless specified otherwise, the calculations are performed using a Perdew-Zunger LDA

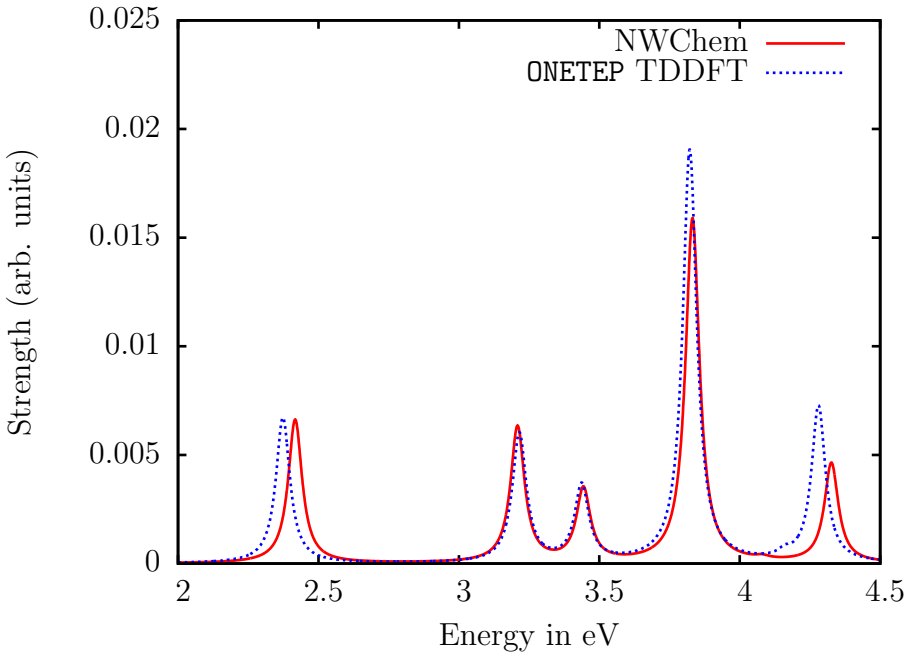


Figure 7.1: Low energy spectrum of a single TiO_2 molecule, as calculated with ONETEP and NWChem. A Lorentzian broadening of 0.03 eV is used in both plots.

exchange-correlation functional [33]. The joint NGWF set is used throughout to represent the unoccupied Kohn–Sham space and the idempotent projector onto the unoccupied subspace is used. The PAW projector sets in all calculations are taken from the ABINIT dataset generated with ATOMPAW [139].

7.2.1 TiO_2 molecule

The first test is performed on a single TiO_2 molecule. Transition metal oxides are notoriously hard to treat with conventional norm-conserving pseudopotentials as they tend to require very hard pseudopotentials and thus a large plane wave basis set to yield converged results [55, 110, 111]. Therefore, the TiO_2 molecule forms an appropriate simple test system. A kinetic energy cutoff of 700 eV is chosen and the box size is given by $30 \times 30 \times 30\, a_0^3$. A minimal set of 4 NGWFs per oxygen and 9 NGWFs per titanium atom is selected for this calculation. Furthermore, an NGWF radius of $13\, a_0$ is chosen for both the conduction and the valence set of NGWFs and the first 4 conduction states are optimised explicitly. The NWChem benchmark calculations are performed for identical atomic positions with the aug-cc-pVTZ Gaussian basis set [129].

The results for the low energy spectrum of the TiO_2 molecule as generated by the 8 lowest singlet states can be found in figure 7.1. One finds that there is in general a very good agreement between the ONETEP and the NWChem results, both for peak positions and oscillator strengths. The maximum energy discrepancy is obtained for the first and

last excitation, which are found to be at an energy of 2.375 eV and 4.282 eV in ONETEP and 2.418 eV and 4.328 eV in NWChem. However given that the ONETEP results are obtained with PAW in periodic boundary conditions while the NWChem results are all-electron results, a discrepancy of less than 50 meV can be considered to be in very good agreement (errors are of the same order of magnitude as typical errors between ONETEP with norm-conserving pseudopotentials and NWChem, as can be seen from section 6.5), especially given that the errors for the other excited states are considerably lower.

While the agreement for the oscillator strength between the PAW and the TDDFT results is very close for the first few excitations, the difference is somewhat more pronounced for the bright state at 3.8 eV and the 8th excitation. However, in general, a very good agreement between the ONETEP and the all-electron results can be observed for this simple example of a transition metal oxide.

7.2.2 Nickel tetracarbonyl

Nickel tetracarbonyl ($Ni(CO)_4$), a small tetrahedral molecule that has four carbon monoxide molecules bound to a central nickel atom, is considered as a second test system. Due to the nickel atom and the four oxygens, this is again an example for a system that is very difficult to treat with conventional norm-conserving pseudopotential approaches. For the ONETEP calculation, a minimal set of 4 NGWFs per carbon and oxygen and 9 NGWFs for the nickel is selected. Calculations are performed using a kinetic energy cutoff of 800 eV and a simulation cell size of $40 \times 40 \times 40\, a_0^3$. An NGWF radius of $13\, a_0$ is used both for the valence and conduction NGWFs and the conduction NGWFs are optimised for the lowest 8 conduction states.

Table 7.1 shows the energies and oscillator strengths of the 12 lowest excitations of $Ni(CO)_4$ as compared to the benchmark calculations performed with NWChem. The TDDFT results have been analysed by a breakdown into individual Kohn-Sham transitions and the ONETEP results have been reordered where necessary to match the order of their corresponding states in NWChem. It can be noted that there is generally a good agreement for states 1-3 and 9-12, both for the excitation energies and oscillator strengths. There is however a significant discrepancy in the results for states 4-8. These states correspond to a triply degenerate state and a doubly degenerate state. In the ONETEP calculation, the order of the two degenerate sets is reversed compared to the NWChem calculation. However, looking at the bandstructure Kohn–Sham energies for both the NWChem calculation and the ONETEP calculation, it is found that this discrepancy is due to a discrepancy in the ordering of the 8 lowest conduction states rather than of a failure of the TDDFT calculation in ONETEP.

Thus, with the exception of states 4 and 5 in the ONETEP calculation, which are found

	ONETEP	NWChem
1	3.265	3.278
2	3.265	3.278
3	3.265	3.278
4	3.616	3.513
5	3.616	3.513
6	3.573	3.555
7	3.573	3.555
8	3.573	3.555
9	3.639(0.0029)	3.604(0.0058)
10	3.639(0.0029)	3.604(0.0058)
11	3.639(0.0029)	3.604 (0.0058)
12	3.683 (0.0115)	3.367(0.0115)

Table 7.1: Results for the 12 lowest singlet excitations of $Ni(CO)_4$ as calculated with ONETEP and NWChem. The ONETEP results have been reordered by comparing their character to that of the NWChem results. Energies are given in eV, oscillator strengths in brackets.

to be about 0.1 eV too high in energy as compared to the NWChem benchmark, the two calculations agree to within a few tenths of meV, which can again be considered a very good agreement given the pathological system treated here. It can thus be concluded that the TDDFT implementation performs as expected and calculations on systems containing transition metals can be performed that yield an overall good agreement with all-electron results obtained from a large Gaussian basis set representation.

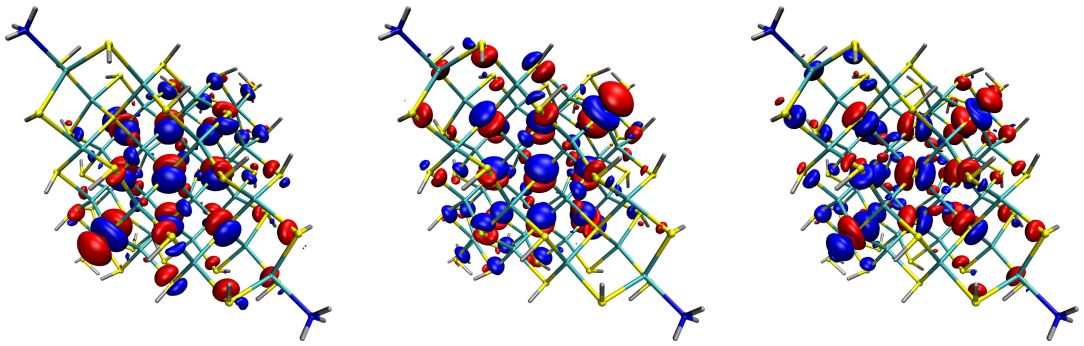


Figure 7.2: Transition densities for the 3-fold degenerate lowest excitation of the CdS nanocrystal at 3 GPa pressure.

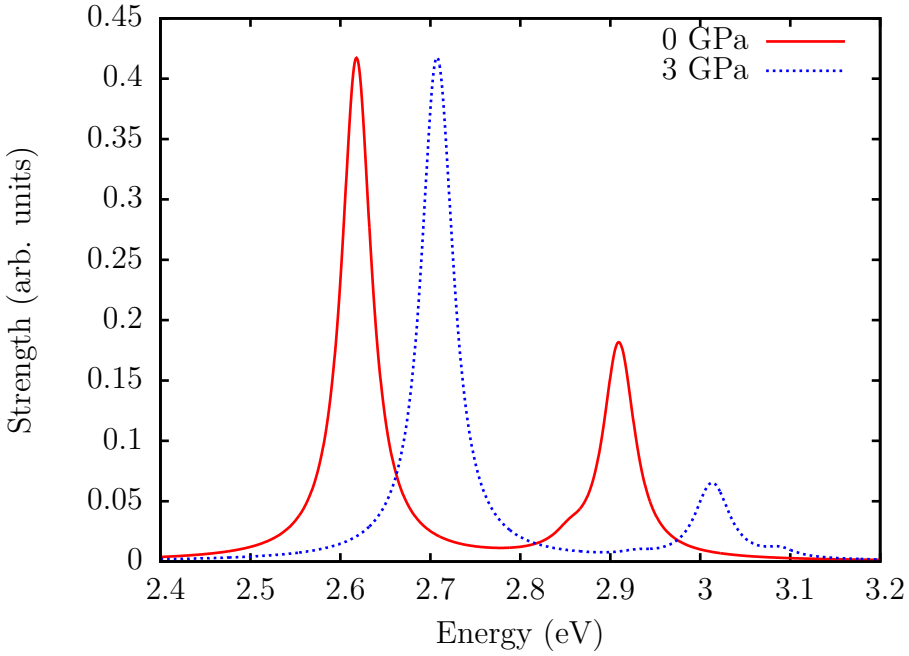


Figure 7.3: Low energy spectrum of a CdS crystal under 0 and 3 GPa pressure, as calculated with ONETEP. A Lorentzian broadening of 0.02 eV is used.

7.2.3 CdS crystals under pressure

Now a more realistic, physically motivated system of a CdS nanoparticle under pressure is considered. Since applying pressure on a nanocrystal lowers the volume of the system, quantum confinement effects tend to increase the Kohn–Sham gap which leads to an increase of the optical gap of the system. The sensitivity of the absorption spectrum of the crystal to pressure leads to a number of interesting potential applications as nanoscale stress sensors [140]. Given that the exact response of the optical properties of a nanocrystal to pressure is very dependent on topology, surface functionalisation and other non-bulk effects, a good theoretical prediction can only be reached if the entire system is treated in a fully *ab initio* way.

Here, a proof-of-principle calculation is performed on the low energy optical spectrum of a small CdS nanocrystal ($Cd_{32}S_{50}H_{48}N_4$) with a surface termination of hydrogens and four NC_3 molecules. Using the method outlined in 5.2.4, the structure of the system is relaxed under 0 and 3 GPa of hydrostatic pressure, yielding two different model systems[38]. The atomic positions from the two structures are then taken to be fixed and a TDDFT calculation is performed in vacuum.

The calculations are performed using a minimal set of 9 NGWFs for the Cd and S atoms, 4 NGWFs for the N atoms and 1 NGWF for the H atoms, for both the conduction

[38]The relaxed structures for the nanocrystal under 0 and 3 GPa of pressure that are used for obtaining the TDDFT results presented here were provided by Niccolo Corsini.

and the valence NGWFs. A cutoff energy of 700 eV is used and the NGWF localisation radius is chosen to be $9.0\,a_0$ and $12.0\,a_0$ for the $\{\phi_\alpha\}$ and $\{\chi_\beta\}$ NGWFs respectively. The conduction NGWFs are optimised explicitly for the 17 lowest conduction sates. All calculations are performed in a $60.5 \times 60.5 \times 60.5\,a_0^3$ simulation cell.

The low energy spectrum of the nanocrystal under 0 and 3 GPa of pressure, as calculated from the lowest 12 excited states, can be found in figure 7.3. It is found that the spectrum of both structures is dominated by a triply degenerate excited state located at an energy of 2.61 eV at 0 GPa pressure and 2.71 eV at 3 GPa pressure. Both structures show that the state couples very strongly with light (see figure 7.2 for a plot of the transition densities corresponding to the peak for the structure at 3 GPa pressure[39]). As expected, quantum confinement effects shift the strong peak of the nanocrystal up by 0.1 eV when a force of 3 GPa is exerted.

The prominent absorption onset for this system is expected to be easily observable under experimental conditions. Here, only a proof-of-principle is provided by performing two calculations at different pressures. It should however be noted that it is feasible to perform the TDDFT calculation shown here for a range of different pressures, and thus characterising the phase space of the crystal transforming under pressure by its absorption onset and strength. These results can then be readily transferred to experimental conditions and a detected change in the spectrum of a given nanocrystal can then be matched to an applied pressure in the system.

While the system chosen here is a bit smaller than the nanocrystals that can be routinely produced experimentally, it is noted that the TDDFT approach developed here is capable of reaching nanocrystals of sizes used directly in experiments [116, 140].

[39] The figure, as well as other plots of atomistic systems and transition densities in later chapters of this work are produced using VMD [141].

Chapter 8

Subsystem TDDFT

In this chapter, some conclusions that can be derived from points raised in Chapter 6 regarding the sparsity structure of the response density matrix are discussed in more detail. These conclusions can be summarised in the following way:

- The linear response TDDFT algorithm presented in this work only retains its linear-scaling properties if $\mathbf{P}^{\{1\}}$ can be made sparse for an excitation of interest.
- The algorithm scales as $\mathcal{O}(N_\omega^2)$ with the number of excitation energies converged. For many systems, the number of excitations within a certain energy window grows quadratically with system size, since the number of possible Kohn–Sham transitions in the energy window grows quadratically. Therefore, converging all excited states within a fixed energy window does have a scaling of $\mathcal{O}(N^4)$ with system size associated with it. However, there are many situations where the number of excitations in a certain energy window increases quadratically but the number of excitations of interest from a physical point of view stays constant. A standard example would be a chromophore in a protein environment or a defect state in a periodic system. In these cases it would be highly desirable to be able to converge targeted excitations in a system[40].
- It has been demonstrated in section 6.8 that the sparsity pattern chosen for $\mathbf{P}^{\{1\}}$ directly influences the character of excitations that can be converged. In the case of a polar nanorod, a spherical cutoff radius on elements of the density matrix is an efficient way to eliminate charge-transfer excitations.

The second point raised is of major importance if the linear-scaling method developed in this work is to be of any practical use for calculating physical properties of previously

[40]It should be pointed out that there are eigenvalue solvers capable of solving for excited states in a narrow energy range without explicitly converging all lower energy excitations of the system and thus provide some form of converging targeted excitations [142]. However, this generally requires the energy of the target excitation to be known to some accuracy and assumes that there are no unwanted eigenstates in the direct neighbourhood of the target state.

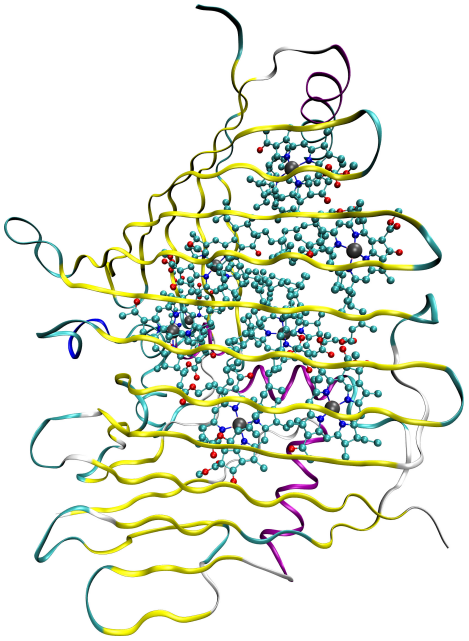


Figure 8.1: Single monomer of the Fenna-Matthews-Olson complex: The chromophores are shown as a ball-and-stick model, while the protein backbone is represented as a ribbon. The Mg atoms of the seven Bchl chromophores are shown in grey.

inaccessible system sizes. The key to bridging the length scales and accessing properties of large-scale systems is to be able to calculate targeted excitations in the system, rather than evaluating the entire low energy spectrum. In this chapter, this problem is addressed and it is demonstrated how to exploit the localised representation of the response density matrix to calculate excitations localised in certain subsystems of a larger system.

8.1 Sparsity of the response density matrix revisited: Subsystems

In order to recognise the power of a subsystem TDDFT approach, it is helpful to first consider a physically motivated example. Here the focus is on the Fenna-Matthews-Olson (FMO) complex, a water-soluble pigment-protein complex (PPC) of green sulphur bacteria (see figure 8.1, the atomic positions are taken from X-ray spectroscopy data obtained in [135]). The PPC consists of three monomers, each containing seven bacteriochlorophyll molecules (Bchls) that are the optically active pigments of the system. The FMO-complex works as a linker complex, funnelling excitons from antenna complexes to reaction centres with a quantum efficiency of close to 100 %. Figuring out the reasons for the highly efficient exciton transport in these structures has been an area of intense research over the last couple of years [143, 144].

The comparatively simple structure and the existence of high quality X-ray crystal-

lographic data of atomic positions make the FMO complex an ideal candidate for studies with computational methods. In the computational approaches used, one normally tries to calculate the site energies of excitons localised on individual Bchls and compute the coupling of individual localised excitons to each other. This data obtained from *ab initio* calculations can then be used to parameterise a model Hamiltonian to study the exciton dynamics of the system [145, 146]. However, even this comparatively small pigment-protein complex is normally too large to be studied fully quantum mechanically by conventional methods, given that a single monomer containing the 7 BChls and the full protein environment consists of roughly 10,000 atoms. Thus most computational studies of the FMO complex rely on some form of mixed quantum mechanical and classical (QM/MM) technique, where a chromophore and some of its surroundings are treated quantum mechanically, while the electrostatic effects of the extended protein environment is represented in form of classical point charges. The problem with this approach is that the influence of the protein environment on the chromophores is often found to be relatively subtle and of a similar order of magnitude to the errors introduced through the QM/MM approach [147, 152][41].

In order to eliminate any potential errors originating from a classical representation of the electrostatic interactions between a chromophore and its surroundings, one would like to treat the entire system on the same footing by performing a DFT and TDDFT calculation including the full FMO complex. However, it is known that the excitations of interest must to some extent be localised on the chromophores, thus representing them with a standard truncated response density matrix is very inefficient. A response density matrix truncated with a spherical cutoff suppresses long range charge transfer style excitations but otherwise allows for excitations delocalised over the entire system, while an excitation localised onto one of the chromophores of the FMO complex should be representable by a much sparser response density matrix.

Thus ideally, one would like to treat the chromophore and its immediate surrounding as a subsystem embedded in the larger protein environment and only compute excitations within that subsystem. In the following section, an algorithm is presented that is capable of systematically converging an excitation confined to a subsystem of a larger system while treating the effects of the entire environment fully quantum mechanically, thus making full use of the sparsity of the response density matrix and allowing for very efficient TDDFT calculations.

[41] As an example, it has been demonstrated that a QM region of 723 atoms or almost half the entire system is necessary to converge the excitation energies on the chromophore in the photoactive yellow protein in a QM/MM approach [153].

8.2 The algorithm

Consider a subsystem **A** consisting of all atoms $\{A_i\}$ of some larger system **U**. While the ground state and conduction state calculations are performed in the usual way for the entire system, the TDDFT excitations should be constrained to be fully localised inside the subsystem **A**. This can be achieved by considering a truncated response density matrix. Labelling the sets of conduction and valence NGWFs of atom A_i as $\{\chi_\alpha^{A_i}\}$ and $\{\phi_\beta^{A_i}\}$, the appropriate response density matrix has the form:

$$\rho^{\{1\}}(\mathbf{r},\mathbf{r}') = \sum_{B_i,C_j} \left[\chi_\alpha^{B_i}(\mathbf{r}) P^{\{1\}\alpha\beta}_{\{B_i,C_j\}} \phi_\beta^{C_j}(\mathbf{r}')\right] \tag{8.1}$$

with

$$P^{\{1\}\alpha\beta}_{\{B_i,C_j\}} = 0 \quad \text{for} \quad B_i \cup C_j \nsubseteq \mathbf{A} \tag{8.2}$$

In the above definition, the block density matrix consisting of all the matrix elements corresponding to NGWFs centered on atoms in **A** is fully dense. In most applications, systems are considered where the spatial extent of the subsystem is small, so that an internal truncation of the block density matrix is negligible. However for large regions **A**, one can combine the above truncation pattern with the usual spherical cutoff radius on the density matrix, causing the block density matrix of the subsystem to become sparse as well.

While the above sparsity structure cannot by definition represent any excitation with contributions from NGWFs centered outside the region **A**, this does not in itself guarantee that the TDDFT algorithm can yield a meaningful lowest excitation of the system, given that $\mathbf{P}^{\{1\}}$ is required to fulfil the invariance constraint specified in equation (6.15) in order to be a valid response density matrix. Fulfilling the constraint while simultaneously preserving the variational properties of the TDDFT algorithm is a key challenge in the subsystem approach.

8.2.1 Invariance constraint

For the invariance constraint of the response density matrix, it is required that $\mathbf{P}^{\{1\}'} = \mathbf{P}^{\{1\}}$, where

$$\mathbf{P}^{\{1\}'} = \mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}, \tag{8.3}$$

the response density matrix projected into the conduction and valence subspace. However, for any sparse $\mathbf{P}^{\{1\}}$, $\mathbf{P}^{\{1\}'}$ is clearly less sparse than $\mathbf{P}^{\{1\}}$, meaning that the relationship can only hold approximately. For a conventionally truncated density matrix like the one considered in section 6.8, the violation of the invariance relationship is generally small

and can be corrected for by iteratively applying the projection in equation (8.3). While the iterative projection causes the TDDFT algorithm to become non-variational, in the GaAs nanorod scenario discussed in section 6.8, the algorithm is found to be well behaved enough to guarantee a good level of convergence for all matrix truncations considered.

However, when considering a response density matrix constrained to a subsystem **A**, one is normally interested in a scenario where the region **A** is much smaller than the entire system **U**. This causes $\mathbf{P}^{\{1\}}$ to be much more sparse than $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{c\}}$ and, by extension, the difference in sparsity between $\mathbf{P}^{\{1\}'}$ and $\mathbf{P}^{\{1\}}$ to be much more problematic, since the amount of violation of the invariance constraint in each conjugate gradient iteration is considerably larger than in the conventional case. The non-variational nature of the algorithm causes the conjugate gradients method to break down and a desired convergence tolerance of the excitation cannot be reached.

The problem of keeping the constraint of the invariance relation while introducing sparsity in the density matrix is to a certain extent analogous with the one encountered in the ground state problem. Thus, just as proposed in the ground state problem, one could use the positive semidefinite norm $Q[\mathbf{P}^{\{1\}}]$ defined in equation (6.30) as a penalty functional, and minimising

$$\Omega_{\text{pen}} = \Omega[\{\mathbf{P}_i^{\{1\}}\}] + \alpha \sum_i^{N_\omega} Q[\mathbf{P}_i^{\{1\}}] \tag{8.4}$$

This approach has exactly the same downside as in ground state DFT, in that it only converges to the correct $\{\mathbf{P}^{\{1\}}\}$ for $\alpha \to \infty$, but has the advantage of being fully variational. One could also use a functional equivalent to the Kohn functional in ground state DFT, by considering $\sqrt{Q[\mathbf{P}_i^{\{1\}}]}$ instead of $Q[\mathbf{P}_i^{\{1\}}]$. This again leads to the correct minimum at the cost that the algorithm is now non-variational again due to the branch point of the square root. Instead, an analogous approach to the ground state DFT LNV method (see section 3.3.5) for TDDFT is introduced in this work.

8.2.2 The auxiliary density matrix

In analogy to the LNV algorithm in ground state DFT, the auxiliary response density matrix $\mathbf{L}^{\{1\}}$ is introduced such that

$$\mathbf{P}^{\{1\}} = \mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{L}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}. \tag{8.5}$$

It is now required that $\mathbf{L}^{\{1\}}$ has a sparsity pattern with matrix elements only non-zero for NGWFs on atoms in subsystem **A** as defined in section 8.2, while $\mathbf{P}^{\{1\}}$ is now considerably less sparse than $\mathbf{L}^{\{1\}}$. To the degree that the sparse density matrices $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{v\}}$ fulfil

their respective idempotency constraints, $\mathbf{P}^{\{1\}}$ fulfils the invariance constraint of equation (6.15). Treating Ω as a functional of $\mathbf{L}^{\{1\}}$ and finding its derivative with respect to the auxiliary density matrix, one finds that the matrix-vector operation $\mathbf{q}^{\chi\phi}$ can be written as

$$\begin{aligned}\mathbf{q}^{\chi\phi} &= \mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{P}^{\{c\}}\mathbf{H}^{\chi}\mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{L}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}} \\ &\quad -\mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{L}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\mathbf{H}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}} \\ &\quad +\mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{P}^{\{c\}}\mathbf{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}\left[\mathbf{P}^{\{c\}}\mathbf{S}^{\chi}\mathbf{L}^{\{1\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}}\right]\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\mathbf{P}^{\{v\}} \end{aligned} \tag{8.6}$$

in terms of $\mathbf{L}^{\{1\}}$. One can now make the approximation that the amount to which the sparse ground state and conduction density matrices violate their idempotency constraints is negligible, which should be the case at the end of any well converged ground state and conduction calculation in ONETEP. Then $\mathbf{q}^{\chi\phi}$ reduces to

$$\mathbf{q}^{\chi\phi} = \mathbf{P}^{\{c\}}\mathbf{H}^{\chi}\mathbf{P}^{\{1\}} - \mathbf{P}^{\{1\}}\mathbf{H}^{\phi}\mathbf{P}^{\{v\}} + \mathbf{P}^{\{c\}}\mathbf{V}_{\mathrm{SCF}}^{\{1\}\chi\phi}\left[\mathbf{P}^{\{1\}}\right]\mathbf{P}^{\{v\}}, \tag{8.7}$$

which is the same familiar form introduced in chapter 6, with the difference that $\mathbf{P}^{\{1\}}$ is now defined via the auxiliary density matrix $\mathbf{L}^{\{1\}}$.

The full TDDFT algorithm then follows a very similar structure to the standard algorithm outlined in chapter 6. One starts with a random guess for $\mathbf{L}^{\{1\}}$ and uses it to construct $\mathbf{P}^{\{1\}}$. The TDDFT gradient is calculated and a conjugate search direction is constructed from $\mathbf{q}^{\chi\phi}$. Once an ideal step size is found, the step direction is used to update $\mathbf{L}^{\{1\}}$ and a new $\mathbf{P}^{\{1\}}$ is constructed. While the use of the auxiliary response density matrix comes at a computational cost since $\mathbf{P}^{\{1\}}$ is significantly less sparse than $\mathbf{L}^{\{1\}}$, it guarantees that the minimisation of Ω is variational and that $\mathbf{P}^{\{1\}}$ fulfils the invariance constraint at every single conjugate gradient step.

The subsystem TDDFT algorithm outlined here then enables the convergence of the lowest TDDFT excitations localised within a certain subsystem of a larger system. Note that the Kohn–Sham Hamiltonian as well as the ground state and conduction density matrices are defined for the entire system. Thus, the only approximation entering the subsystem TDDFT calculation is the sparsity pattern of $\mathbf{L}^{\{1\}}$ defined through the spatial extent of the subsystem $\mathbf{A}$. Following the physically motivated example of the FMO complex, the localised excitation on a single chromophore can be converged by systematically increasing the subsystem region to include parts of the protein environment.

8.2.3 Exciton couplings between subsystems

Consider the case where there is not a single region **A** of interest, but a number of regions **A**, **B**, **C** etc. An example for this has been considered with the FMO complex in section 8.1, but there is a large number of photoactive biological systems that generally have more than one chromophore. While it is possible to converge the excitations localised in each individual region using the above technique, this does not represent a good approximation to the excitations of the full system if there is non-negligible coupling between the subsystems.

If the coupling between subsystems is small, which is a requirement for the subsystem approach to be valid to begin with, its net effect on excitation energies and oscillator strengths can be reintroduced perturbatively. Consider two subsystems **A** and **B** with local response density matrices $\left\{\mathbf{P}^{\{1\}}_{\mathbf{A},i}; i = 1, ...N^{\mathbf{A}}_{\omega}\right\}$ and $\left\{\mathbf{P}^{\{1\}}_{\mathbf{B},j}; j = 1, ...N^{\mathbf{B}}_{\omega}\right\}$. It is then assumed that any global excitation of the coupled subsystems can be written as a linear combination of $\{\mathbf{P}^{\{1\}}_{\mathbf{A},i}\}$ and $\{\mathbf{P}^{\{1\}}_{\mathbf{B},j}\}$. Constructing the two-particle TDDFT Hamiltonian $\mathbf{H}^{[2p]}_{\mathrm{TDA}}$ in the representation of the combined solution space $\{\mathbf{P}^{\{1\}}_{\mathbf{A},i}\} \oplus \{\mathbf{P}^{\{1\}}_{\mathbf{B},j}\}$ the following effective eigenvalue equation is obtained:

$$\mathbf{H}^{[2p]}_{\mathbf{A},\mathbf{B}}\mathbf{u} = \omega\mathbf{S}_{\mathbf{A},\mathbf{B}}\mathbf{u} \tag{8.8}$$

where the block matrices $\mathbf{H}^{[2p]}_{\mathbf{A},\mathbf{B}}$ and $\mathbf{S}_{\mathbf{A},\mathbf{B}}$ are being given by

$$\mathbf{H}^{[2p]}_{\mathbf{A},\mathbf{B}} = \begin{pmatrix} \omega_{\mathbf{AA}} & \Omega_{\mathbf{AB}} \\ \Omega_{\mathbf{AB}} & \omega_{\mathbf{BB}} \end{pmatrix} \tag{8.9}$$

with matrix elements

$$\omega_{\mathbf{A},i\mathbf{A},j} = \delta_{ij}\omega_{\mathbf{A}_j} \tag{8.10}$$

$$\Omega_{\mathbf{A},i\mathbf{B},j} = \mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}_{\mathbf{A},i}\mathbf{S}^{\chi}\mathbf{q}^{\chi\phi}_{\mathbf{B},j}\mathbf{S}^{\phi}\right] \tag{8.11}$$

and

$$\mathbf{S}_{\mathbf{A},\mathbf{B}} = \begin{pmatrix} \mathbf{I}_{\mathbf{AA}} & \mathbf{S}_{\mathbf{AB}} \\ \mathbf{S}_{\mathbf{AB}} & \mathbf{I}_{\mathbf{BB}} \end{pmatrix} \tag{8.12}$$

with matrix elements

$$I_{\mathbf{A},i\mathbf{A},j} = \delta_{ij} \tag{8.13}$$

$$S_{\mathbf{A},i\mathbf{B},j} = \mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}_{\mathbf{A},i}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}_{\mathbf{B},j}\mathbf{S}^{\phi}\right]. \tag{8.14}$$

Just like in the subspace diagonalisation of the conventional TDDFT algorithm, $\mathbf{H}^{[2p]}_{\mathbf{A},\mathbf{B}}$

and $\mathbf{S_{A,B}}$ can be trivially constructed from $\left\{\mathbf{P}^{\{1\}}_{\mathbf{A},i}\right\}$, $\left\{\mathbf{P}^{\{1\}}_{\mathbf{B},j}\right\}$ and their corresponding $\left\{\mathbf{q}^{\chi\phi}_{\mathbf{A},i}\right\}$ and $\left\{\mathbf{q}^{\chi\phi}_{\mathbf{B},i}\right\}$. Solving the eigenvalue problem in equation (8.8) then yields all excited states of the global system, while the matrix of eigenvectors $\mathbf{u}$ can be used to construct the full transition density matrices $\left\{\mathbf{P}^{\{1\}}_{\mathbf{tot},k}; k = 1, ...(N^{\mathbf{A}}_{\omega} + N^{\mathbf{B}}_{\omega})\right\}$ from the transition density matrices of the individual subsystems $\mathbf{A}$ and $\mathbf{B}$. Expanding the formalism from two coupled subsystems to an arbitrary number of coupled systems is trivially done.

8.2.4 Scaling considerations

Just like in the conventional linear-scaling TDDFT algorithm, the subspace diagonalisation of equation (8.8) scales as the cube of the size of the total subspace $\left\{\{\mathbf{P}^{\{1\}}_{\mathbf{A},i}\} \oplus \{\mathbf{P}^{\{1\}}_{\mathbf{B},j}\} \oplus\{\mathbf{P}^{\{1\}}_{\mathbf{C},k}\} \oplus ...\right\}$ considered in the calculation. However, given that one is usually only interested in a small number of excitations per subsystem, the subspace diagonalisation is never a bottleneck in practical calculations.

Now consider the computational cost of evaluating the low energy excitation spectrum of a system $\mathbf{U}$ containing N_{sub} subsystems of interest. Performing the ground state DFT and the conduction optimisation on the entire system $\mathbf{U}$ scales linearly with system size. To obtain the spectrum of the entire system it is necessary to perform N_{sub} individual subsystem TDDFT calculations, converging a fixed number of states in each case. However, note that the sparsity pattern of the auxiliary response density matrix $\mathbf{L}^{\{1\}}$ of a subsystem dictates that the number on nonzero elements it contains must stay constant, rather than increase linearly, with system size. Given that the number of states of interest contained in each individual subsystem is also constant with system size, obtaining these states for an individual subsystem scales as $\mathcal{O}(1)$ with system size of the global system $\mathbf{U}$. The number of individual subsystems of interest N_{sub} can be assumed to grow linearly with system size, therefore obtaining all transition density matrices for all localised excitations of interest for all subsystems scales as $\mathcal{O}(N)$ with system size. As discussed above, the computational subspace diagonalisation that has to be performed when coupling the individual subsystems is negligible for most practical calculations.

One can therefore conclude that calculating the excitation spectrum for some low energy window of a system that can be divided into individual subsystems with well defined localised excitations scales fully linearly with the size of that system. If the separation into subsystems is possible, the subsystem TDDFT approach described here avoids the scaling problems stated at the beginning of this chapter and allows for truly $\mathcal{O}(N)$ calculations of excitation spectra. Clearly, this division is not always possible in practice. While it is expected to yield very good results for chromophores in light harvesting biological systems or organic crystals, it is not likely to work in situations

where it is not clear *a priori* where the excitation localises or in systems where well localised excitations simply do not exist. However, in the remainder of this work it will be demonstrated that the subsystem TDDFT method is very powerful in a number of interesting realistic systems.

8.3 Comparison to the FDEc method

The method of converging localised excitations in different regions **A**, **B** ... and then coupling them in a post-processing step by assuming that global transition density matrices can be written as linear combinations of transition density matrices of the individual subsystems is analogous to the coupled frozen-density-embedding TDDFT (FDEc-TDDFT) approach introduced by Neugebauer *et al.* [148–150]. However, their approach shows a number of differences, mainly in how the ground state calculation is treated. This leads to some differences in the form of the coupled subsystem TDDFT eigenvalue equations, which will be briefly summarised here without going into too much detail on the FDEc method.

In the subsystem TDDFT approach presented here, the effects of the embedding environment are fully included without the need of any approximation, while in FDEc they are included by calculating an approximate embedding potential. The FDEc method divides the total density of the system in two parts, that of the subsystem and that of the embedding environment. While it is trivial to write down a Hohenberg-Kohn density functional for the total energy of a system partitioned in this way, the Kohn–Sham treatment is difficult as the Kohn–Sham single particle kinetic energy of the total system is not the sum of the single particle kinetic energies of its constituent parts. This means that one has to add an extra term for the non-additive kinetic energy to the energy functional. Since the exact functional form of the non-additive kinetic energy term is generally unknown, it has to be approximated in practical calculations and thus introduces a further approximation compared to the conventional Kohn–Sham scheme.

The advantage of the FDEc method is that only a small part of the system has to be treated quantum mechanically at any given time during the ground-state and TDDFT calculation, while the effect of the rest of the system is included through the embedding potential. In the TDDFT method introduced here, the Hamiltonian and the ground state density matrix are calculated for the entire system. This removes any additional approximations in the form of a non-additive kinetic energy functional but means that a larger system has to be treated fully quantum mechanically. Here, the advantage of the linear-scaling DFT method becomes apparent as it allows for the full treatment of very large systems without the need of a fragment approach.

After the TDDFT excitation energies are found for individual subsystems, the coupling of the subsystems in the FDEc method follows a similar approach as in the subsystem TDDFT method presented here, in that the full TDDFT equation is rewritten in the subspace of eigenvectors of the individual local excitations of all the subsystems. However, the block matrix elements $\Omega_{\mathbf{AB}}$ coupling the two subsystems are different in the two approaches. In the FDEc approach, the coupling originates from $\mathbf{V}_{\mathrm{SCF}}^{\{1\}}$, which contains an additional term due to the second derivative of the non-additive kinetic energy with respect to the density of the system. In the subsystem TDDFT approach presented here, the kinetic energy term is naturally missing. Instead the coupling elements also contain contributions from the diagonal part of Kohn–Sham eigenvalue differences:

$$\Omega^{\mathrm{diag}}_{\mathbf{A},i\mathbf{B},j} = \mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}_{\mathbf{A},i}\mathbf{S}^{\chi}\mathbf{P}^{\{c\}}\mathbf{H}^{\chi}\mathbf{P}^{\{1\}}_{\mathbf{B},j}\mathbf{S}^{\phi}\right] - \mathrm{Tr}\left[\mathbf{P}^{\{1\}\dagger}_{\mathbf{A},i}\mathbf{S}^{\chi}\mathbf{P}^{\{1\}}_{\mathbf{B},j}\mathbf{H}^{\phi}\mathbf{P}^{\{v\}}\mathbf{S}^{\phi}\right] \tag{8.15}$$

This term is only strictly vanishing if $\mathbf{P}^{\{1\}}_{\mathbf{A},i}$ does not contain any fractions of a Kohn–Sham transition that are also present in $\mathbf{P}^{\{1\}}_{\mathbf{B},j}$, that is to say if $\mathbf{P}^{\{1\}}_{\mathbf{A},i}$ and $\mathbf{P}^{\{1\}}_{\mathbf{A},i}$ are orthogonal and span entirely different subspaces. This is not guaranteed to be the case in the subsystem approach presented here, although the overlap between $\mathbf{P}^{\{1\}}_{\mathbf{A},i}$ and $\mathbf{P}^{\{1\}}_{\mathbf{A},i}$ can be assumed to be very small for well separated and thus well defined subsystems. This diagonal term is readily available in the subsystem TDDFT approach presented here because the Hamiltonians $\mathbf{H}^{\chi}$ and $\mathbf{H}^{\phi}$ are evaluated for the entire system and thus $\mathbf{P}^{\{1\}}_{\mathbf{A},i}$ and $\mathbf{P}^{\{1\}}_{\mathbf{B},j}$ can both be represented in the same total space of all possible Kohn–Sham transitions of the total system.

It can be summarised that the FDEc and other density embedding approaches to TDDFT achieve a similar goal to the method proposed here, in that they both perform a global TDDFT calculation by breaking the total system into different subsystems with well localised excitations that can be coupled in a post-processing step involving a single subspace diagonalisation. The method introduced here is conceptually very simple, as the convergence of excitations localised in subsystems is achieved purely by truncating the auxiliary response density matrix. Thus the accuracy of the method is uniquely defined by a single parameter, the size of the region of the subsystem of interest. This approach relies heavily on the fact that it is possible to perform a ground state DFT calculation of the entire system in linear-scaling effort within the `ONETEP` method. The FDEc method never requires the treatment of the full system in the ground state or excited state calculations but has the disadvantage of introducing additional approximations.

Both the FDEc method and the subsystem method introduced here make the inherent assumptions that there exist meaningful localised excitations in the system and that the response density matrices of the entire system can be written as a linear combination of the

density matrices of the individual subsystems. As mentioned earlier, the first assumption breaks down as soon as it is not possible to define a meaningful, physically motivated region in the full system where excitations might have a local character. The second assumption breaks down as soon as the coupling between the two subsystems becomes strong. This breakdown leads back to the first assumption, because in the limit of strong coupling between subsystems **A** and **B**, it is no longer meaningful to regard them as well defined individual subsystems.

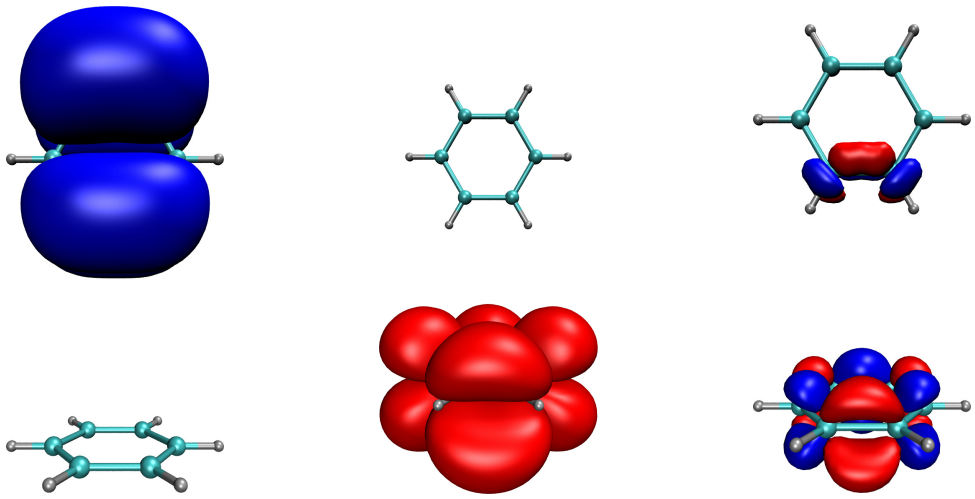
8.4 Benchmark tests

In order to test the subsystem algorithm, a number of benchmark tests are performed. The test calculations are relatively simple examples that are used to establish the validity of both the convergence of excitations localised in certain parts of larger systems and the coupling of excitations between individual subsystems. In the next chapter, the algorithm established here is used on realistic systems of real research interest. All calculations presented here are performed using norm-conserving pseudopotentials and the ALDA functional in the Perdew-Zunger parameterisation is used unless specified otherwise. Furthermore, one NGWF per H atom and four NGWFs per C and O atom is used throughout and the conduction space manifold is represented by the joint NGWF set $\{\theta_\alpha\}$ and the projection operator onto the entire unoccupied subspace.

8.4.1 Benzene

As a first test system, a benzene dimer in a T-shape configuration is considered in order to test the capability of converging targeted excitations with the density matrix truncation. For this purpose each benzene molecule is treated as an individual subsystem, causing the $\mathbf{L}^{\{1\}}$ matrices to have 75 % of all elements set to zero. As the system is too small to make use of any matrix truncations for the conduction and valence density matrices, $\mathbf{P}^{\{1\}}$ is still fully dense for both subsystems. The T-shape configuration is chosen to ensure that the dipole moments of excitations localised on the two different benzene molecules are orthogonal to each other, and it is therefore expected that the coupling between the two subsystems is vanishingly small for most low energy excitations. Thus one expects to find low energy excited states that are well localised on a single benzene molecule and that are thus well represented by the subsystem approach.

Calculations are performed on a box size of $40 \times 40 \times 40\, a_0^3$, using a kinetic energy cutoff of 800 eV. A centre of mass separation of $12.35\, a_0$ is chosen for the two benzene molecules in the T-shape configuration. The localisation radii are set to $12\, a_0$ and $10\, a_0$ for the $\{\chi_\alpha\}$ and $\{\phi_\beta\}$ NGWF representations respectively and the PBE functional [39]



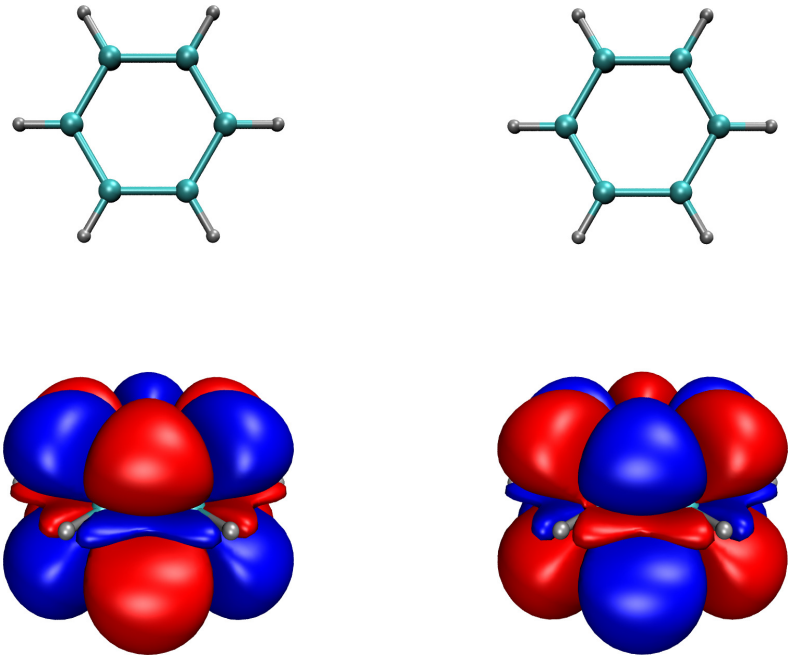
(a) Hole Density (b) Electron Density (c) Transition Density

Figure 8.2: Electron, hole and transition density for the first excitation of a T-shaped benzene dimer for a fully dense response density matrix

is used for all calculations. First, a benchmark calculation is performed where the entire system is treated explicitly, no truncation on the response density matrix is used and the 16 lowest singlet states are converged. Then, the lowest 8 localised singlet excitations are computed for both of the benzenes, first treating one benzene as the subsystem of interest and then the other. As a final step, the two sets of response density matrices from the localised excitations of both benzenes are used to construct the coupled subspace eigenvalue problem. This subspace matrix is then diagonalised to find the 16 eigenstates of the coupled system.

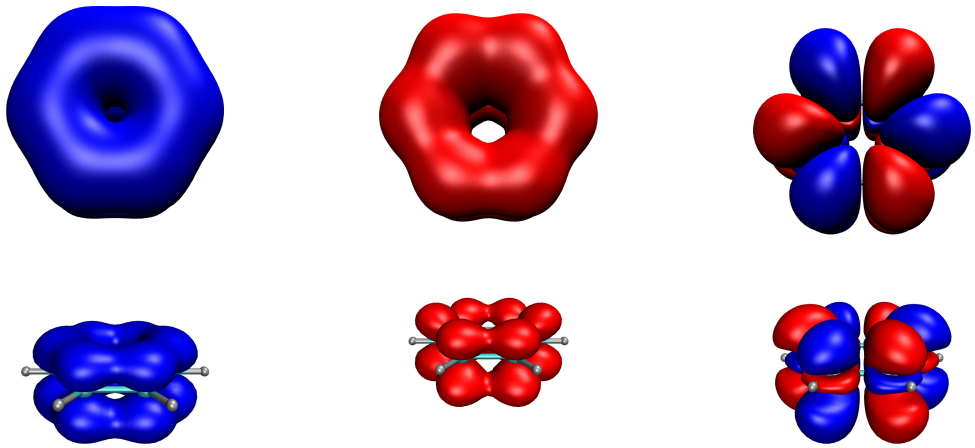
Since a semi-local exchange correlation functional is used, it is again expected that the lowest excitation of the system corresponds to an unphysical charge transfer state from one benzene to the other. As can be seen from figure 8.2, this is exactly what can be observed. The lowest excitation is found to have an energy of 4.892 eV and a vanishing oscillator strength. Figure 8.2 shows that the hole is entirely localised on the upper benzene, while the electron is fully localised on the lower benzene, with the transition density being spread over both of them. A breakdown of the TDDFT transition into Kohn–Sham states confirms that the excitation is fully composed of an excitation from the HOMO of the upper benzene to the LUMO of the lower benzene, as expected.

In fact, the first eight excitations of the system treated with a fully dense $\mathbf{L}^{\{1\}}$ correspond to unphysical charge transfer states that all originate from the failure of the semi-local exchange correlation kernel to capture non-local exchange effects. However, the 9th excitation corresponds to an excited state localised on a single benzene. As seen



(a) 9th excitation full system (b) 1st excitation truncated system

Figure 8.3: Transition density of the 9th excitation of the entire system compared to the transition density of the first excitation of a system where $\mathbf{L}^{\{1\}}$ only has non-zero elements on the lower benzene.



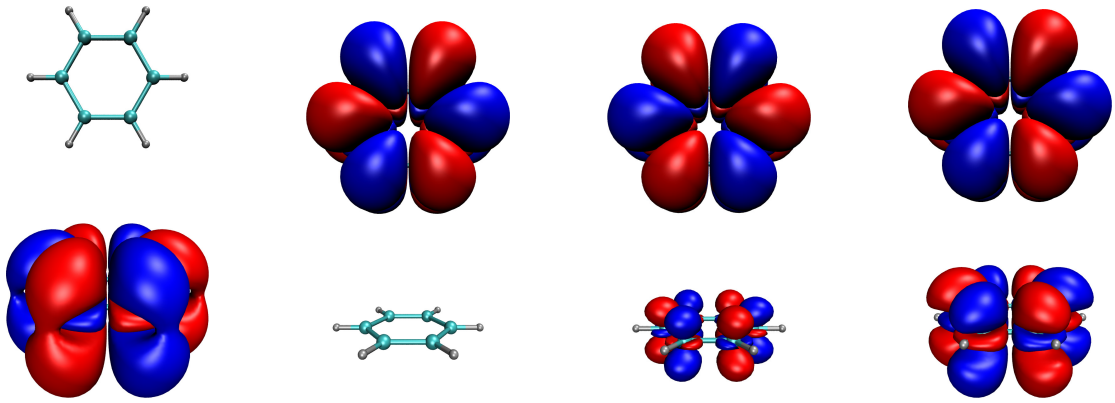
(a) Hole Density (b) Electron Density (c) Transition Density

Figure 8.4: Electron, hole and transition density for the 11th excitation of the full system.

in figure 8.3, the transition densities for the 9th excited state of the dense system and the 1st excited state of the system where $\mathbf{L}^{\{1\}}$ is constrained to have non-zero elements for the lower benzene only are virtually identical, apart from a difference in the sign of the transition vector. However, the sign of the transition vector is arbitrary and has no effect on the excitation energy and thus the two different state vectors indeed describe the same excitation. This is confirmed by a breakdown of the transitions into their respective Kohn–Sham transtions.

The excitation energy obtained with the fully dense $\mathbf{L}^{\{1\}}$ is 5.1950 eV, while the truncated response density matrix yields a marginally higher value of 5.1953 eV. Thus the error induced due to constraining the auxiliary transition density matrix is of the order of 0.3 meV. Here, in this simple test system, one can already observe the strength of the subsystem TDDFT approach. Just as with the radial cutoff used for the GaAs nanorod in section 6.8, constraining the response density matrix allows for the direct convergence of the localised excitations well described by (semi)-local exchange correlation functionals, while removing any charge transfer excitations from the system for which the choice of exchange correlation functional must necessary fail. This strongly limits the number of excitations that have to be converged to obtain the excitations of interest, yielding a significant increase in computational efficiency.

However, while the 9th excitation of the full system shows a localised character well described by a single subsystem, this is not true for all excitations of the system that do not have charge-transfer character. Consider the 11th excitation of the system, plotted in figure 8.4. While this excitation is mainly localised on the upper benzene molecule, the electron and the hole clearly spread out over both benzenes. Therefore, the excitation



(a) Transition density benzene 1 (b) Transition density benzene 2 (c) Transition density coupled (d) Transition density whole system

Figure 8.5: Transition density of the 11th excitation of the full system as compared with the transition density of the same excitation obtained in the coupled subsystem approach. The two main localised transitions on benzene 1 and benzene 2 that are contributing to the coupled excitation are also plotted.

does not have a pure charge transfer character from one benzene to the other. However, given that the transition density is spread out over both benzenes, it cannot be obtained from a single $\mathbf{L}^{\{1\}}$ localised on one of the benzenes.

The main contribution to the 11th excitation comes from the upper benzene, so it is to be hoped that the full excitation can be reproduced by taking a localised excitation on the upper benzene and mixing in some contributions from the lower benzene. As can be seen from figure 8.5, this is precisely what is found. The excitation is dominated by the second excitation of the subsystem of the upper benzene. This excitation couples to the second, third and fourth excitation of the subsystem of the lower benzene (in the figure, only the 2nd excitation of the lower system is plotted), with the resulting transition density being similar to the one of the 11th excitation of the full system.

The main differences between the two transition densities is that there is less weight on the lower benzene in the coupled approach. Furthermore, while the sign of the transition density of the upper benzene is reversed, that of the lower benzene is not when compared to the full solution. In general, the coupling between the two subsystems is found to be very weak, due to the T-shape configuration of the benzene dimer that does not allow for strong dipole interactions between the two benzenes. Thus the excitation in question is already well described by a localised transition density on the upper benzene and shifting some weight to the lower benzene does not significantly alter the excitation energy. The excitation energy for the coupled system is found to be 6.2685 eV, in good agreement with the 6.2643 eV obtained for the full system. Thus the error of the subsystem approach is

of the order of a few meVs in this example.

It can be concluded that it is possible to converge localised excited states in the T-shape dimer of benzene that agree very well to excitations obtained for the whole system. The excitations of the full system tend to show some more spread in the transition density that is not fully reproduced by the coupled subsystem. However, due to the very weak coupling between subsystems, the excitation energies of the more localised subsystem excitations are in excellent agreement with the energies for the full system. Thus the calculations presented here can be taken as a demonstration that it is indeed possible to converge targeted excitations localised in some subsystem of a larger system with the algorithm outlined above.

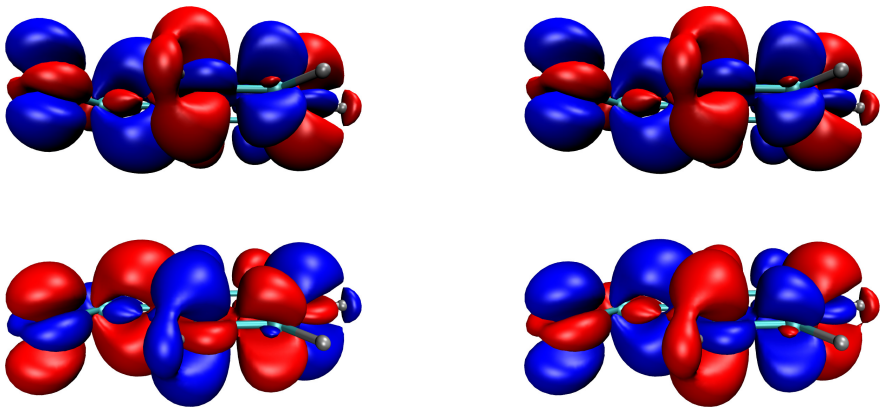
8.4.2 Benzaldehyde

While the example of a T-shape benzene dimer demonstrates some of the strengths of the subsystem approach, such as the removal of any unwanted badly-represented charge-transfer excitations from the subspace of allowed solutions, it does not provide an adequate test of the coupling of the subsystems, since the coupling of the benzenes in the T-shape configuration is so weak.

For this purpose, a calculation on a dimer of benzaldehyde (C_6H_5CHO) is performed. Rather than choosing a T-shape configuration like in the case of benzene, the molecules are stacked on top of each other so that they are aligned, thus maximising the interaction between the two systems. In such a configuration, one can then observe exciton splitting: the two benzaldehyde molecules each have a set of localised excitons associated with them, which are, due to the symmetry of the problem, necessarily degenerate in energy. When these excitons couple to form the total excitation of the entire system, they can do so in two different ways. Either, their transition densities can align such that the transition dipole moments align or they can oppose each other such that the transition dipole moments cancel one another. The two different configurations can be seen in figure 8.6 for a chosen excited state of benzaldehyde.

The interaction between the two excitons causes the dipole-opposed configuration to have a lowered energy and vanishing oscillator strength, while the dipole aligned configuration shows an enhanced oscillator strength and an increased excitation energy. Thus the degenerate exciton splits into two different excited states, purely due to the coupling between the systems. This coupling is expected to be well reproduced by the subsystem TDDFT approach.

Thus a subsystem TDDFT calculation of the benzaldeyhde dimer is performed at different dimer separations and the results for a chosen exciton pair are compared to the results obtained when the entire system is treated explicitly. The simulation cell size is



(a) Dipole-opposed configuration (b) Dipole-aligned configuration

Figure 8.6: Exciton splitting in a benzaldehyde dimer. The exciton splits into a dipole aligned configuration with enhanced oscillator strength and higher excitation energy and a dipole opposed configuration with vanishing oscillator strength and lowered excitation energy.

chosen to be $40 \times 40 \times 40\, a_0^3$ and a kinetic energy cutoff of 1200 eV is used. The Perdew-Zunger exchange-correlation functional [33] is used and an NGWF localisation radius of $10\, a_0$ and $12\, a_0$ is chosen for the valence and conduction NGWFs respectively. For the subsystem TDDFT approach, the four lowest excitations localised on each benzaldehyde are converged and then coupled, while for the treatment of the entire system, the lowest 14 states are converged explicitly. It is found that even for this small system there is a significant saving in computational cost when the subsystem approach is considered, as converging the two sets of localised excitons and coupling them is found to only require about 55 % of the time taken to converge the 14 lowest states of the system with a dense response density matrix.

The exciton splitting between the degenerate excitons corresponding to the third transition in the individual benzaldehydes, a bright state with an oscillator strength of 0.0166 in the isolated system, is then considered. Just as expected, for large dimer separation, this state splits into one bright state with an enhanced oscillator strength of 0.033 and a dark state that is lower in energy.

Figure 8.7 shows the energy of the two excited states in question, obtained from both the coupled subsystem approach and the full TDDFT calculation on the entire system, for different dimer separations. As can be clearly seen, for large enough dimer separations, the excitation energies obtained with the coupled subsystem approach agree perfectly with the full TDDFT calculation. As the two benzaldehyde molecules get closer together, the exciton splitting between the excited states grows larger, as one would expect from

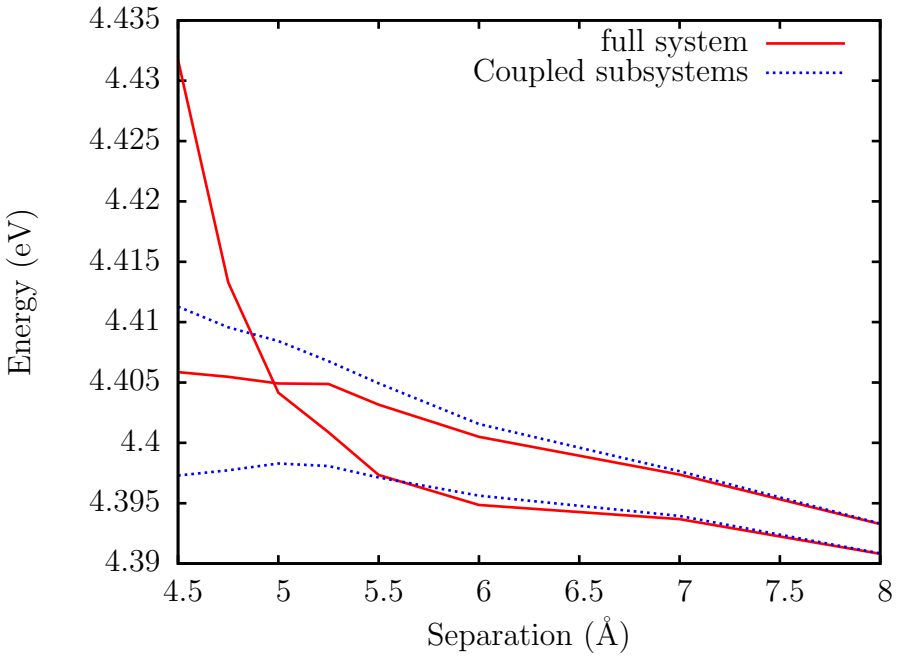


Figure 8.7: Exciton splitting in Benzaldehyde plotted against dimer separation for both the full system and the coupled subsystem approach.

the increased interaction strength between the excitations. However, up to a separation of around 5.5 Å, the subsystem calculation still yields a good agreement with the full TDDFT approach.

This good agreement breaks down for smaller separations. While the subsystem TDDFT approach simply predicts an increased splitting between the two excitons, the full TDDFT approach shows that a crossing between the excited state energy surfaces of the two excitons occurs at a dimer separation of around 5 Å. Here, the breakdown of the transition density matrix into Kohn–Sham transitions is used to confirm that a crossing of excited state potential energy surfaces is indeed occurring, rather than an avoided crossing. Naturally this crossing fundamentally changes the character of the excitations such that they are no longer well represented by the simple model of two interacting localised excitons, leading to a breakdown of the subsystem TDDFT approach.

In conclusion, it is found that the subsystem TDDFT approach is indeed capable of describing the coupling between excited states of interacting subsystems to great accuracy, as long as these subsystems are not interacting strongly. In the limit of strong interaction, the excitations can no longer be written as a linear combination of localised excitations of individual subsystems and the subsystem approximation must necessarily break down. Note that in this limit it is no longer physically meaningful to treat the system as consisting of well defined localised subsystems that are weakly interacting and there is no alternative to treating the entire system explicitly. However, it should be pointed out that a number

of the systems discussed as potential applications to this method, such that chromophores in biological systems and organic crystals, are known to interact only weakly and thus the subsystem approach is expected to be valid. Some of these systems are considered explicitly in the next chapter.

Chapter 9

Large-scale applications

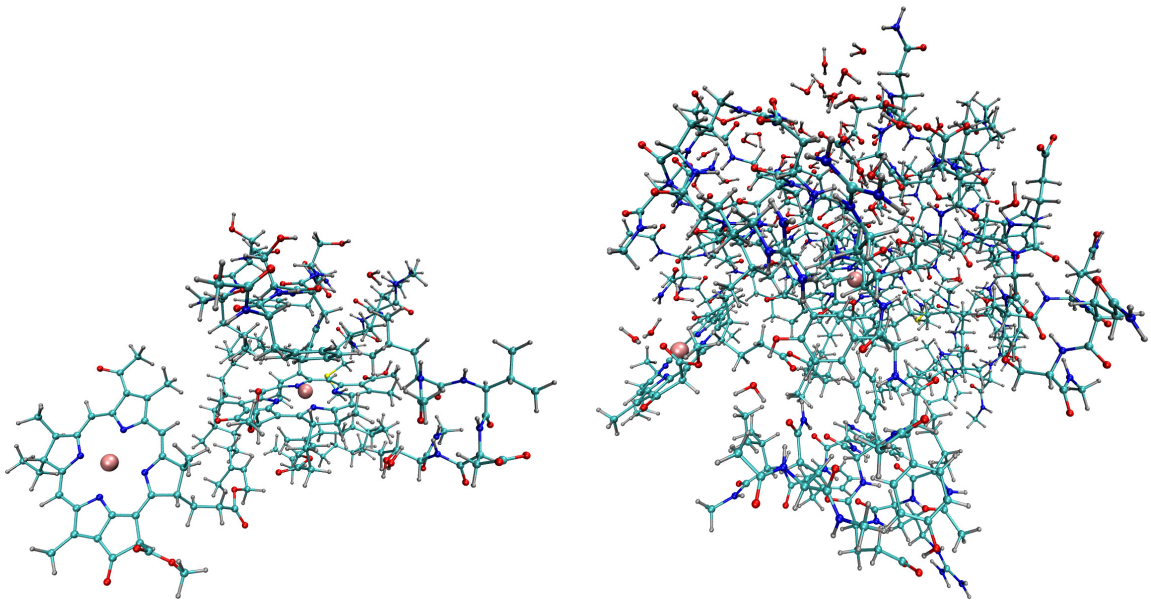
To test the methods developed in the last couple of chapters, a number of realistic large scale applications are focused on. Two different problems are considered in detail, the first one of which, the study of environmental effects on the excitons localised on bacteriochlorophyll (Bchl) sites in the Fenna-Matthews-Olson complex, has already been briefly introduced in the last chapter. As a second problem, the singlet and triplet excitations of pentacene-derived dopant molecules are studied when substituted into an infinite *p*-terphenyl organic crystal.

9.1 The Fenna-Matthews-Olson (FMO) complex

In order to assess the influence of the protein environment on localised excitation energies in the Fenna-Matthews-Olson (FMO) complex, two model systems from the full complex are designed. For the first model, the Bchl site 1 of the X-ray structure obtained in [135] is taken and all atoms of the protein environment that are within a radius of 10 Å centered on the Mg atom of site 1 are included, as well as the full nearest neighbour Bchl molecule (corresponding to Bchl2 following the original labelling by Fenna and Matthews [151]). For the second model, protein environment up to a radius of 15 Å is included and again, Bchl2 is included completely as well. Thus, two structures for the Bchl molecule of site 1 are obtained, with different amounts of the protein environment included, where the smaller model corresponds to a system size of 562 atoms, while the larger model contains 1646 atoms. Both models are shown in figure 9.1[42].

The 10 Å and 15 Å structures of the two models correspond exactly to the ones discussed in [152], where it is found that the smaller system of 10 Å cutoff radius is sufficient to converge the Qy transition of the system constructed from pure Kohn–Sham

[42]The optimised structures of both models of Bchl1, equivalent to those used in [152] were provided by Daniel Cole



(a) 562 atoms (b) 1646 atoms

Figure 9.1: Two different structures used in the FMO calculation: The 10 Å cutoff radius corresponds to a system with 562 atoms, while the 15 Å cutoff radius corresponds to a system of 1646 atoms.

transitions[43]. Here, the discussion is extended to the case where the exciton energies are obtained with TDDFT rather than just taken to be the Kohn–Sham eigenvalue differences of Kohn–Sham states localised on the Bchl1 site.

Calculations are performed in a $101.25^3\, a_0^3$ simulation cell for the small model and a $123.75^3\, a_0^3$ cell for the larger model and an implicit solvent model (see section 5.2.3) with a static dielectric constant of $\epsilon_{\text{static}} = 80$ is used throughout in order to remove any spurious protein states in the HOMO-LUMO gap due to insufficient electrostatic screening at the surfaces of the systems. The dynamic response of the solvent to the excitations is modelled by choosing $\epsilon_{\text{dynamic}} = 2$, approximately corresponding to the optical response of water. The kinetic energy cutoff used in both models is 1020 eV and a cutoff radius of $10.0\, a_0$ is used for all NGWFs. All calculations in this section are performed using norm-conserving pseudopotentials and the LDA functional. For both models, the 16 lowest conduction states are explicitly optimised and the joint representation and the projector onto the entire unoccupied subspace is used in the TDDFT calculations. Since both models are of moderate size, an explicit density matrix truncation of $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{c\}}$ is deemed unnecessary and both the conduction and the valence DFT calculations are performed using fully dense density matrices.

[43]However, to converge Qy transitions of other Bchl sites it is often found that the larger system of 15 Å is needed. See the supplementary information of reference [152] for more details.

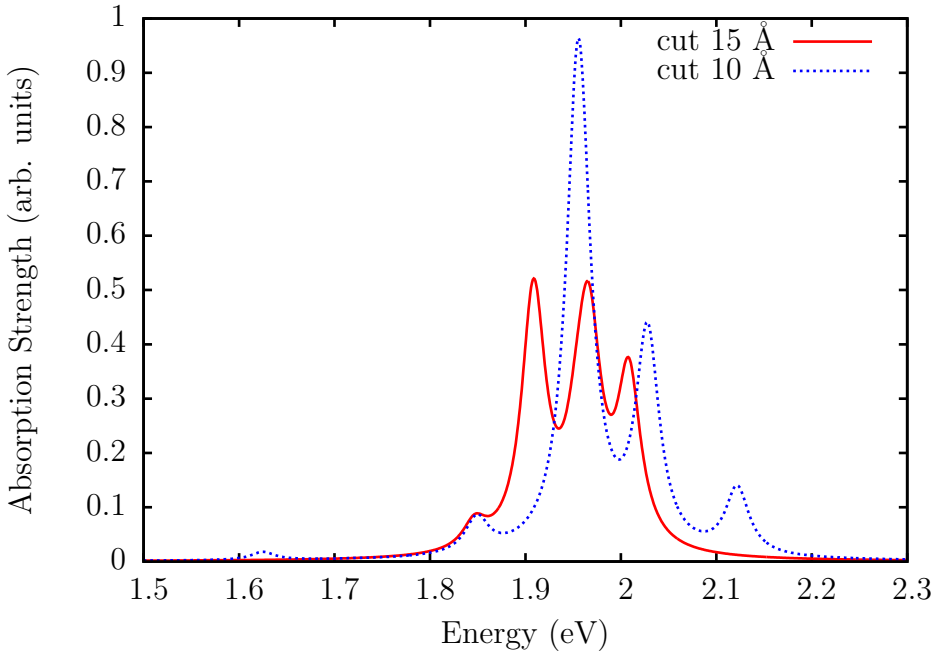
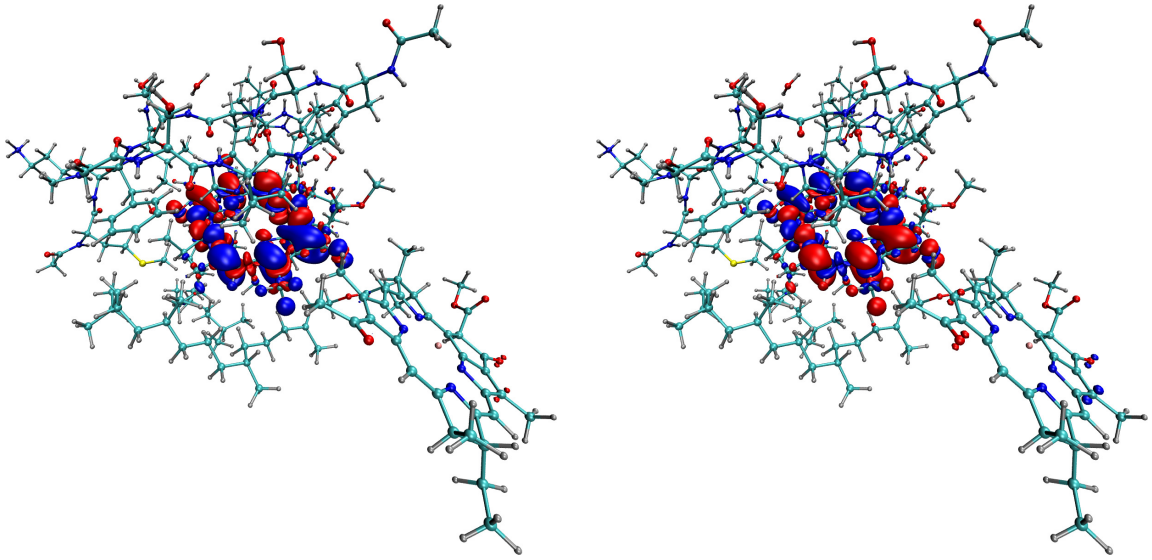


Figure 9.2: Low energy spectrum of the 10 Å cutoff system and the 15 Å cutoff system, both calculated using a fully dense $\mathbf{P}^{\{1\}}$. A Lorentzian broadening of 15 meV is used in both spectra.

In a first step, a full TDDFT calculation is performed on the two models, using fully dense response matrices $\mathbf{P}^{\{1\}}$. Due to the presence of the two Bchl sites in the models and the use of a local exchange-correlation kernel, the treatment of the systems with fully dense response matrices yields a number of unphysical low energy charge transfer states between the Bchl sites. Since these states are dark, they do not show up in the spectrum but do nevertheless have to be converged in order to obtain all excitations in some low energy window. Furthermore, even though the systems are placed in an implicit solvent, there are a number of spurious occupied and unoccupied protein states close to the HOMO and LUMO states respectively that are caused by electrostatic effects due to the way the models are cut out of the larger FMO system. These spurious protein states cause the low energy spectrum to contain a further number of charge transfer states, all of which are dark as well. For this reason, in order to calculate the spectrum in an energy range between 1 and 2.2 eV, it is necessary to explicitly calculate the 14 lowest excitations of the 10 Å model and the lowest 34 excitations of the 15 Å model.

The low energy spectrum of the two models can be found in figure 9.2. Note that the spectrum for the smaller model is dominated by four peaks of different intensity, while the larger model shows an extended feature between 1.90 and 2.05 eV that is made up of three distinct peaks of similar intensity. Analysing the spectrum of the small model, it is found that the first and the second distinct peak with excitation energies 1.850 eV

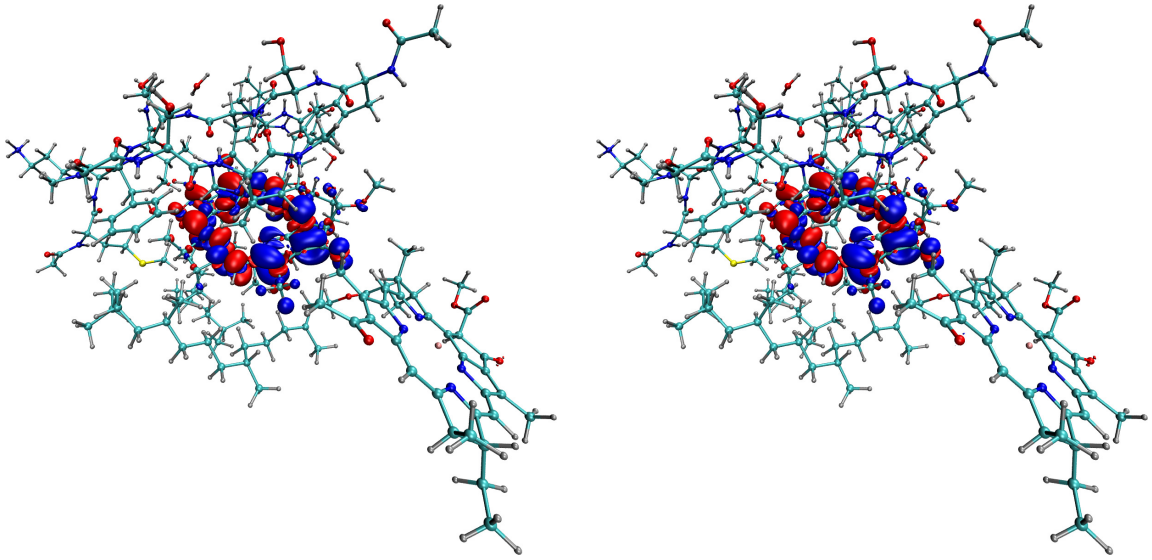


(a) coupled subsystems (b) full system

Figure 9.3: Transition density of the S_1 excitation localised on Bchl1 for the 10 Å cutoff system, obtained by coupling the two subsystems and by solving for the entire system with a dense $\mathbf{P}^{\{1\}}$.

and 1.956 eV respectively are associated with excitons localised on the central Bchl1 of interest, while the third and fourth peaks at 2.029 eV and 2.123 eV are associated with Bchl2. Here, the strongest peak in the spectrum corresponds to the Qy transition of Bcl1, while the first small peak is the Qx transition. Both Qy and Qx transitions show non-negligible coupling to Kohn-Sham transitions associated with Bcl2. For the larger model, the first peak is found to correspond to the Qx transition of Bchl1 and, with an energy of 1.848 eV, is found at almost exactly the same energy as the Qx transition in the smaller model. The second peak in the 15 Å model corresponds to the Qy transition of Bchl1 and is located at an energy of 1.909 eV. The fourth peak in the spectrum is associated with Bchl2 and is located at 2.009 eV. The Bchl2 excitation corresponding to the fourth peak in the smaller model is absent, most likely due to the fact that not enough excitations were calculated in the larger system to obtain it. The third peak in the large model is an interesting case: its character is shared between Bchl1 and Bchl2 and it is entirely absent from the smaller model.

For the purpose of calculating site energies in the Fenna-Matthews-Olson complex, we are mainly interested in how the Qx and the Qy transitions of Bchl site 1 behave with the amount of protein environment considered explicitly. Thus we focus on the two lowest excitations of both models. It can be noted that the Qx transition associated with Bchl1 in the large model is almost exactly reproduced in the smaller model, both in excitation energy and oscillator strength. However, the Qy transition shows significant



(a) coupled subsystems (b) full system

Figure 9.4: Transition density of the S_2 excitation localised on Bchl1 for the 10 Å cutoff system, obtained by coupling the two subsystems and by solving for the entire system with a dense $\mathbf{P}^{\{1\}}$.

differences between the two models. In the larger model, the energy of the transition is lowered by almost 50 meV, decreasing the splitting between Qx and Qy energies by the same amount. Furthermore, the oscillator strength of the excitation is halved compared to the smaller model. These discrepancies can be directly attributed to the difference in environmental screening of the excitation in the two different systems. It can thus be concluded that while the Qx and Qy transition energies derived from simple Kohn-Sham energy differences reported in [152] are converged for the small model of Bchl1, the same is not true for the excitation energies derived from TDDFT. While the test calculation performed here does not allow us to say whether the Qy energy obtained from the 15 Å model is converged with the amount of explicit protein environment included in the calculation, it is clearly evident that the small model fails to provide a converged value for the Qy transition, suggesting that the explicit treatment of large systems is indeed necessary in determining site energies in the Fenna-Matthews-Olson complex.

After having established that an explicit quantum mechanical treatment of a large amount of the protein environment seems to be necessary to correctly converge an excitation localised on a Bchl site in the FMO complex, it is now interesting to consider the question of whether these excitations can be converged with the coupled subsystem approach. It can be seen from the spectrum in Figure 9.2, that the low energy absorption is dominated by the low energy excitations localised on the two Bchls and thus converging the spectrum using a fully dense $\mathbf{P}^{\{1\}}$ is wasteful and leads to a large number of spurious

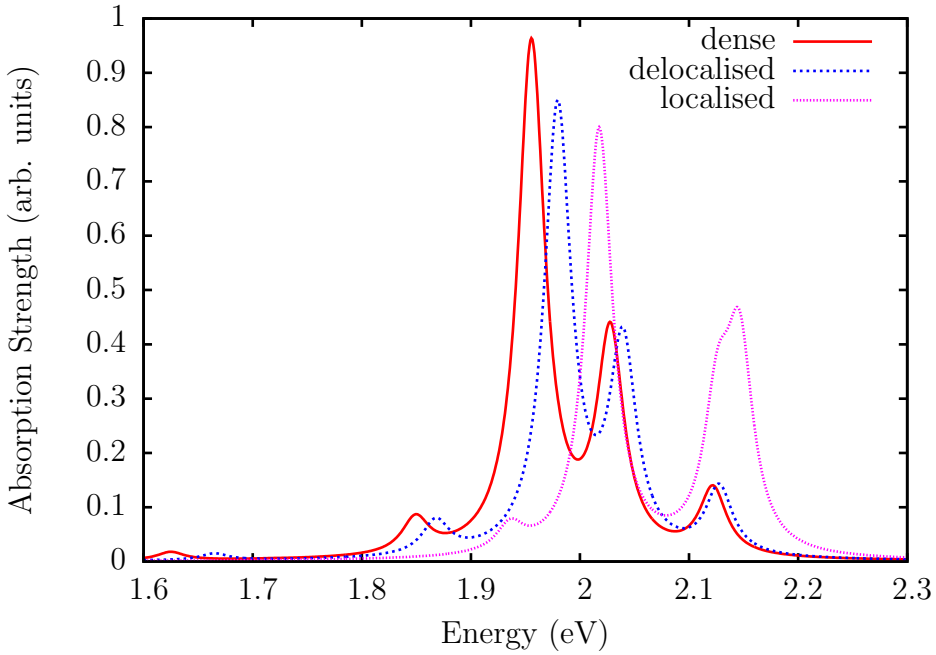
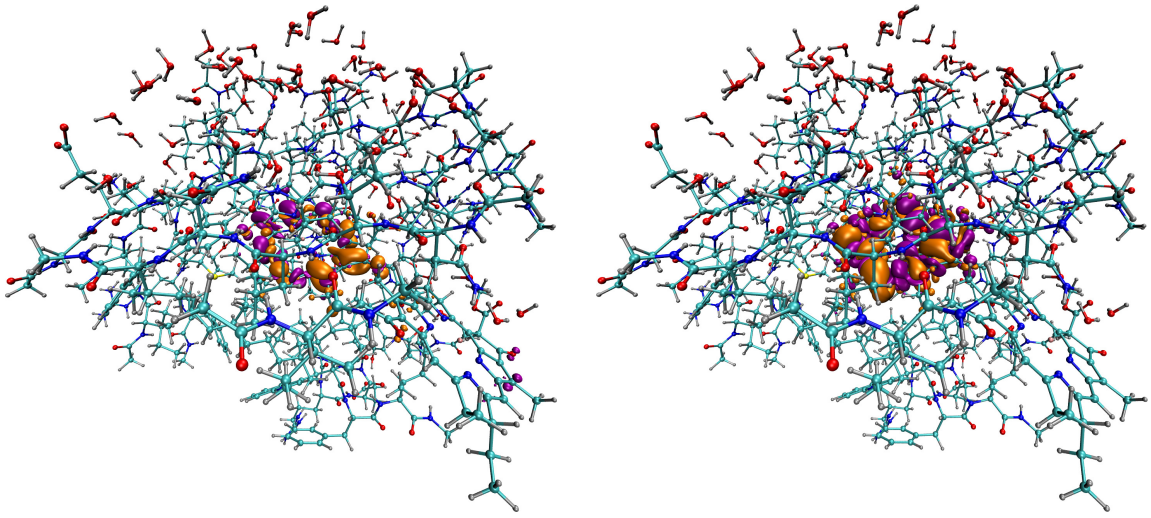


Figure 9.5: Low energy spectrum of the 10 Å cutoff system as calculated with a fully dense $\mathbf{P}^{\{1\}}$ or the subsystem approach. Two different subsystem definitions are used. For the spectrum labelled "fully local", $\mathbf{P}^{\{1\}}$ is restricted onto one of the Bchls only. For the spectrum labelled "delocalised", $\mathbf{P}^{\{1\}}$ contains the protein environment as well. A Lorentzian broadening of 10 meV is used in all spectra.

dark charge transfer states.

In order to establish the validity of the coupled subsystem approach for this realistic system that is considerably more complicated than the model systems treated in section 8.4, the small model is focused on first and two different definitions of the subsystems for the response density matrices are used. In the first, fully localised definition of the subsystems, subsystem **A** and **B** are taken to only contain atoms belonging to one of the Bchls but none of the atoms belonging to the protein environment. In a second, delocalised definition of the subsystems, subsystem **A** contains all atoms of Bchl1 and all protein atoms, but none of the atoms belonging to Bchl2, while subsystem **B** contains all atoms in the system apart from atoms belonging to Bchl1. The lowest few singlet states for each of the subsystem are converged and are then coupled as described in chapter 8. Generally it is found that coupling between the subsystems is weak, with coupling strengths between excitations of the order of a few meV, and that the subsystem approach is thus well defined.

The results for the low energy spectrum of the small model, as calculated with the fully dense approach and the two definitions of subsystems can be found in figure 9.5. It can be seen that when the subsystems are taken to be fully localised on the Bchls, the excitations corresponding to Bchl1 are significantly overestimated, with errors of about



(a) S_1 (b) S_2

Figure 9.6: Transition density of the first and second singlet excitation localised on Bchl1 for the 15 Å cutoff system, where both excitations are obtained by coupling the two subsystems.

0.1 eV, while the discrepancies for Bchl2 are somewhat smaller but the oscillator strength of the transitions is wrong. When the protein environment is included in the subsystem approach, the errors are significantly reduced, with excitation energies being overestimated by about 20 meV as compared to the calculation corresponding to a fully dense $\mathbf{P}^{\{1\}}$. There is a slight discrepancy in oscillator strengths for the Qy transition of Bchl1 in the subsystem approach, which is slightly overestimated. However, in general the agreement between the subsystem method and the full approach is very good, with the advantage that the subsystem TDDFT approach eliminates a number of spurious charge transfer states by design and is much more efficient computationally.

In figures 9.3 and 9.4 the transition density of the two low energy singlet states localised on Bchl1 are plotted as calculated with the delocalised coupled subsystem approach and the full TDDFT approach. As it can be seen, the agreement is very good. Furthermore, note that in 9.3, the Qy transition can be seen to have a small contribution on Bchl2, that is correctly reproduced in the coupled subsystem approach.

Figure 9.7 shows the coupled subsystem approach as applied to the larger model in comparison to the full TDDFT treatment. Here, the delocalised definition of the subsystems is used that does include the protein environment but excludes all spurious charge-transfer states between the Bchls. The errors of the subsystem approach are slightly larger in the large system, with the Qy transition on Bchl1 being overestimated by about 40 meV. This suggests that the subsystem region should be defined more carefully, possibly including some parts of Bchl2 that are very close to Bchl1 in order to get fully converged

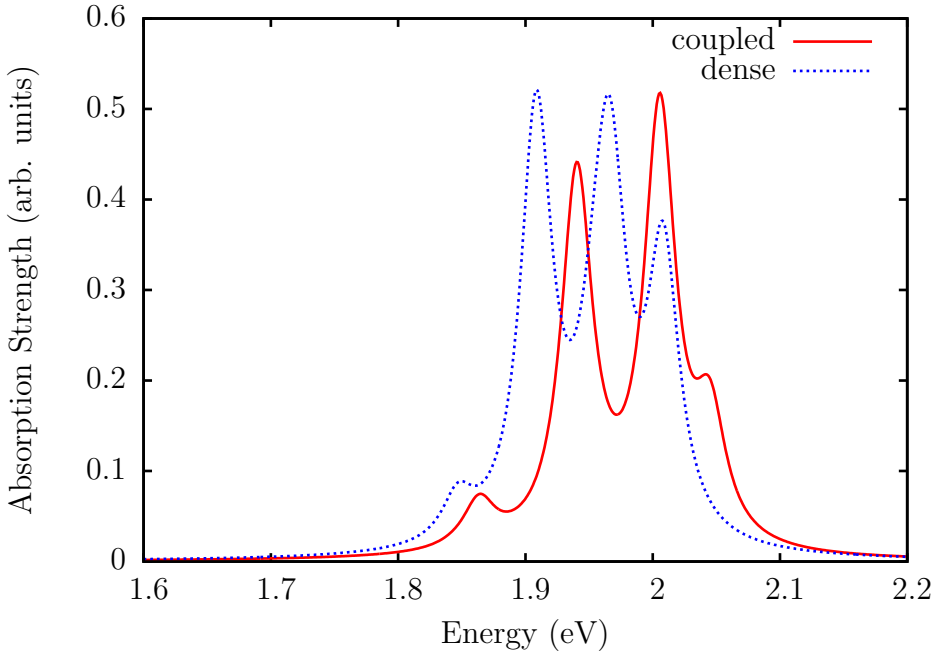


Figure 9.7: Low energy spectrum of the 15 Å cutoff system as calculated with a fully dense $\mathbf{P}^{\{1\}}$ or the subsystem approach. For the coupled subsystem approach $\mathbf{P}^{\{1\}}$ is limited to one of the Bchls and the full protein environment, but excludes the other Bchl. A Lorentzian broadening of 10 meV is used in both spectra.

answers.

When looking at the S_1 and S_2 states as obtained in the coupled subsystem approach for the large model (see figure 9.6) in comparison to the ones obtained for the smaller model, it can be noticed that the transition densities indeed look very similar as expected, but that the S_2 state seems to delocalise considerably more in the larger model.

In conclusion the computational study performed here shows two different effects. First of all, an explicit inclusion of large parts of protein environment into the calculation is crucial to obtain well-converged excitation energies for localised excitons in the FMO complex, making the treatment of very large systems necessary. Secondly, the subsystem TDDFT approach is a viable method to significantly decrease the computational load for large scale calculations of excitation energies on the chromophores in the FMO complex, capable of converging energies to an accuracy of 10-20 meV for the small system and 30-40meV for the larger system. The main features of spectra are well reproduced in the subsystem approach, with oscillator strength in reasonable agreement to those of the full system. Drawing from these conclusions, linear-scaling DFT and linear-scaling subsystem TDDFT can be proposed as a viable method to obtain the low energy optical spectrum of the FMO complex, while it is demonstrated that the explicit treatment of large parts of the protein environment is crucial in order to obtain converged site energies.

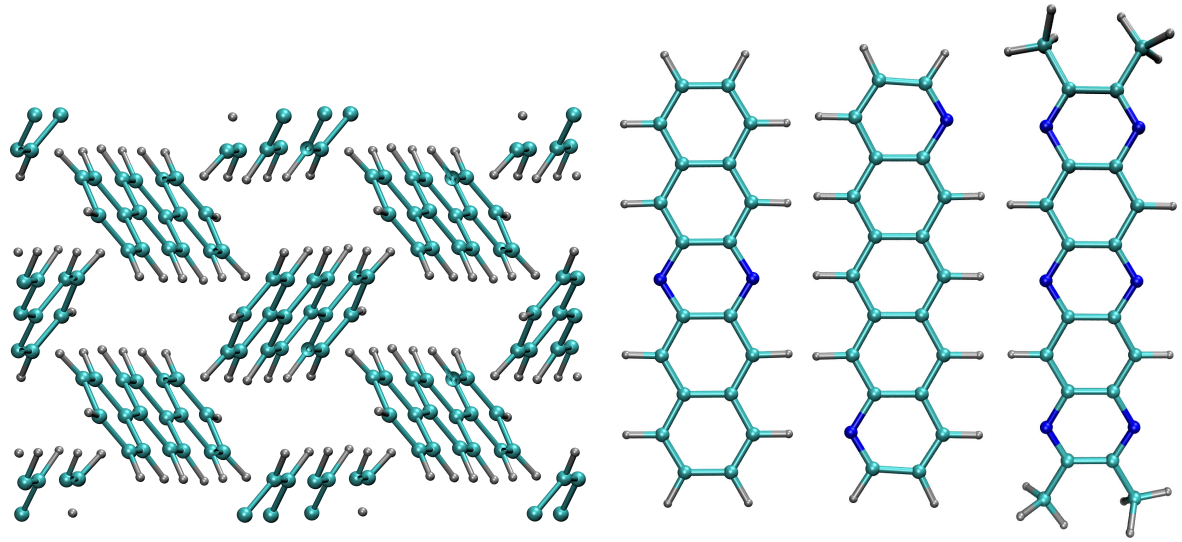
9.2 Exciton delocalisation on molecular dopants in a *p*-terphenyl crystal

The focus is now shifted onto the second large scale application of the TDDFT method developed in this work, which is the study of excitons of pentacene-derived dopant molecules embedded in an infinite *p*-terphenyl crystal. From a theoretical point of view, this type of system is of immense interest as it was demonstrated that pentacene defects in *p*-terphenyl organic crystals can be used to create a zero-field room-temperature maser [154].

The way the room-temperature maser operates is by pumping the first singlet state of the pentacene molecule, which couples strongly to light. The excitation undergoes an intersystem crossing to the first triplet state T_1 via the second triplet state T_2, which is close in energy to S_1. The process is spin-selective and therefore leads to a population inversion between the highest and lowest sublevel of T_1. When this metastable population inversion is triggered to revert through stimulated emission to fully populate the lowest energy sublevel of T_1, the system emits a microwave pulse of about 1.45 GHz, equivalent to the spitting between the highest and lowest sublevel of the triplet state [154].

From the above considerations it becomes clear that the performance of the maser relies on a number of factors. First of all, the pentacene dopant molecule must fit well into the *p*-terphenyl structure, such that its structure is only mildly perturbed and the S_1, T_1 and T_2 states retain a similar character to the case of an isolated pentacene molecule. Second, in order for the intersystem crossing to be viable, the energy difference between the S_1 and the T_2 state must be relatively small so that little energy has to be gained or lost via vibrational modes in order to cross from the singlet to the triplet level. Finally, the T_1 state must fill in a way such that a natural population inversion occurs and after the masing has happened, the T_1 state must relax back to the ground state S_0 at a fast enough rate to allow for continued maser pulses.

The purpose of this study is to investigate some of the above mentioned conditions for pentacene and a number of other proposed pentacene-derived maser molecules embedded in a realistic model of *p*-terphenyl. The fact that no spin-orbit interaction is considered in this work means that the splitting of the T_1 state and the rate of intersystem crossing cannot be predicted as these properties require the evaluation of spin-orbit coupling matrix elements. Furthermore, the standard expected error in TDDFT excitation energies does not allow for determining whether the T_2 and S_1 states line up perfectly in the real molecule, as the TDDFT error for a given triplet state is not necessarily of the same size as for a given singlet state. However, since it is known experimentally that pentacene works as a dopant, it is possible to compare the computational results for different dopants to that of pentacene, given that a pentacene-derived dopant molecule is expected to show



(a) c-DNP (b) o-DNP (c) HNTMP

Figure 9.8: Figure showing the crystal structure of *p*-terphenyl, as well as the relaxed structures of the three molecular defects that are proposed additionally to pentacene. The pentacene derivatives are labelled after the number of nitrogen atoms substituted into the structure. To differentiate between (a) and (b), a label of 'c' for 'central' and 'o' for 'opposite' is introduced to denote the position of the carbon atoms that are replaced.

very similar TDDFT errors as pentacene. Thus, a combined TDDFT and DFT approach allows for a prescreening of potential maser molecules on the basis of structural properties and the alignment of excited states with those obtained for pentacene. It is then possible to determine whether a given molecule is a likely candidate for being used in a room-temperature maser.

All DFT calculations in this section are performed with an ALDA functional [33] and norm-conserving pseudopotentials. A kinetic energy cutoff of 750 eV is used and a localisation radius of $10\,a_0$ is chosen for all NGWFs both in the conduction and the valence NGWF set. No density matrix truncation on $\mathbf{P}^{\{v\}}$ or $\mathbf{P}^{\{c\}}$ is applied throughout this section.

For the computational study, a $4 \times 4 \times 3$ supercell of a *p*-terphenyl crystal is chosen, containing 72 *p*-terphenyl molecules and a total number of 2304 atoms. The position of the carbon atoms, as well as the cell shape and volume are fixed to experimental results obtained from [155], while the hydrogen atoms are added and their positions are optimised at the LDA level. Fixing the carbon molecules to the experimental positions guarantees that a realistic crystal structure is obtained, since DFT cannot produce the correct binding behaviour between the individual molecules at LDA level due to the lack of van der Waals interactions.

	pentacene	c-DNP	o-DNP	HNTMP
av. force	0.108	0.087	0.095	1.575
max. force	0.212	0.135	0.159	9.741

Table 9.1: Average and maximum force acting on the four defect molecules after being placed into the *p*-terphenyl matrix. Forces are given in eV/Å.

Four different dopant molecules are considered, one being pentacene and the other three being pentacene derivatives obtained by replacing different C-H atoms by nitrogen atoms. The pentacene derivatives chosen are 6,13-dinitro-pentacene (c-DNP), 2,9-dinitro-pentacene (o-DNP) and 1,4,6,8,11,13-hexanitro-2,3,9,10-tetramethyl-pentacene (HNTMP) (see figure 9.8 for relaxed structures in vacuum for these molecules). Together with pentacene, these three candidate molecules are then to be substituted into the *p*-terphenyl crystal, replacing one of the *p*-terphenyl molecules. While the supercell of *p*-terphenyl is too small to reach realistic dopant concentrations of 10^{-4} for the pentacene as reported in [154], it is considered to be large enough to provide a realistic representation of the environmental effects experienced by the embedded dopant molecule.

In a first step the geometries of all dopant molecules are relaxed in vacuum and the low energy triplet and singlet states are calculated. These benchmark calculations can then be used to identify the influence of the *p*-terphenyl crystal on excitation energies of the dopant molecules. The vacuum structures of the molecules are then taken and substituted into the *p*-terphenyl crystal for one of the *p*-terphenyl molecules. The substitution is performed such that the dopant molecule is oriented in the same way as the *p*-terphenyl it is substituted for and that their centres of mass coincide. A ground state calculation with fixed atomic positions is then performed and the maximum and average forces on the substituted molecules are calculated.

The results of the force calculations can be found in table 9.1. As it can be seen, the pentacene and the two DNP structures show relatively small forces when placed into the *p*-terphenyl structure. Maximum forces on the pentacene are slightly higher than for the two DNP structures, but average forces are around 0.1 eV/Å. This suggests that all three dopant molecules fit very well into the *p*-terphenyl structure and are likely to cause only minimal perturbations to their immediate surroundings. The c-DNP dopant shows the best fit, which can be explained by the fact that equilibrium C-N bond lengths are shorter than C-C bonds, meaning that the length of the c-DNP is closer to that of *p*-terphenyl than the length of o-DNP or pentacene. The HNTMP molecule however shows different characteristics, with a maximum force that is almost two orders of magnitude larger than the maximum force on the c-DNP molecule and an average force 15 times larger than the average force on the other molecules. The high forces are mainly due to the methyl groups

attached to the end of the pentacene derivative, making the dopant molecule HNTMP considerably longer than the *p*-terphenyl it is substituted for. Thus it is expected that the HNTMP molecule can only be fitted into the *p*-terphenyl crystal by causing significant distortions to the surrounding crystal structure.

Following the above considerations, in order to obtain realistic structures for the defect molecules, a geometry optimisation of them placed inside the *p*-terphenyl structure is performed. For the pentacene and the DNP structures, the surrounding atoms are kept fixed and only the dopant molecule is relaxed. Since the forces are already very small to begin with, only a few iterations are needed to reduce the maximum force to values lower than 0.1 eV/Å. For the HNTMP structure, a geometry optimisation of the dopant molecule only in a fixed *p*-terphenyl crystal fails to converge, suggesting that it is indeed necessary to perturb the surrounding crystal. Thus for the HNTMP molecule, a full geometry optimisation on the entire system is performed and all atoms are allowed to relax, yielding significant perturbations on nearest neighbour *p*-terphenyl molecules. The maximum change in position for any atom in the relaxed crystal compared to the atomic positions in the fixed crystal is found to be 1.04 Å, with an average change of 0.12 Å per atom.

After having obtained realistic structures for all four dopant molecules substituted into the *p*-terphenyl crystal, a TDDFT calculation is performed on each of the four systems. Here, the main interest is in singlet and triplet excitons that are localised on the dopant molecule and retain a similar character as in vacuum. However, in many situations excitons on molecules in organic crystals are known to show a degree of delocalisation to neighbouring molecules [156, 157], thus using a truncation on $\mathbf{P}^{\{1\}}$ to force a complete localisation of excitons on the defect molecule is not expected to yield realistic results. On the other hand, the local nature of the ALDA functional is likely to delocalise the exciton much too strongly and couple it to charge transfer states to all *p*-terphenyl molecules in the crystal in the case where no truncation on $\mathbf{P}^{\{1\}}$ is used. Therefore, calculations are performed using a $\mathbf{P}^{\{1\}}$ that is non-zero only for the defect molecule and the six nearest neighbour *p*-terphenyls. Thus, delocalisation of the exciton to nearest neighbour *p*-terphenyl molecules is allowed, while any long-range charge transfer excitations are suppressed.

However, while long range charge transfer states are effectively excluded from the subset of allowed solutions it is still not guaranteed that the obtained excitations have the character of excitons localised on the defect molecule with only small amounts of delocalisation. Table 9.2 shows the Kohn–Sham energies of the HOMO, LUMO an HOMO-1 states of the four different systems, as measured with respect to the LUMO+1 state. Note that the HOMO and LUMO states correspond to Kohn–Sham states localised on

	pentacene	c-DNP	o-DNP	HNTMP
HOMO-1	-2.726	-2.682	-2.683	-2.394
HOMO	-1.775	-2.294	-2.066	-2.233
LUMO	-0.627	-1.228	-0.904	-0.972
HOMO-LUMO gap	1.148	1.066	1.162	1.261

Table 9.2: Kohn–Sham energies of the HOMO-1, HOMO and LUMO states for all 4 crystals with embedded defect molecules, as measured in eV with respect to the LUMO+1 state.

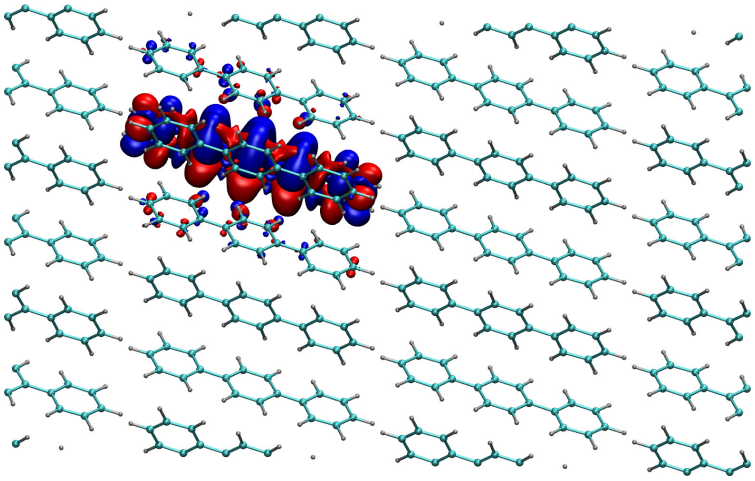
	pentacene vacuum	pentacene *p*-terphenyl	o-DNP vacuum	o-DNP *p*-terphenyl
T_1	0.946	0.965	0.955	0.980
T_2	2.125	2.123	2.122	2.113
S_1	1.856 (0.0431)	1.803 (0.0679)	1.854 (0.0430)	1.748 (0.0658)

Table 9.3: T_1, T_2 and S_1 states for pentacene and o-DNP, both in vacuum and in the *p*-terphenyl crystal. Energies are given in eVs, oscillator strengths in brackets.

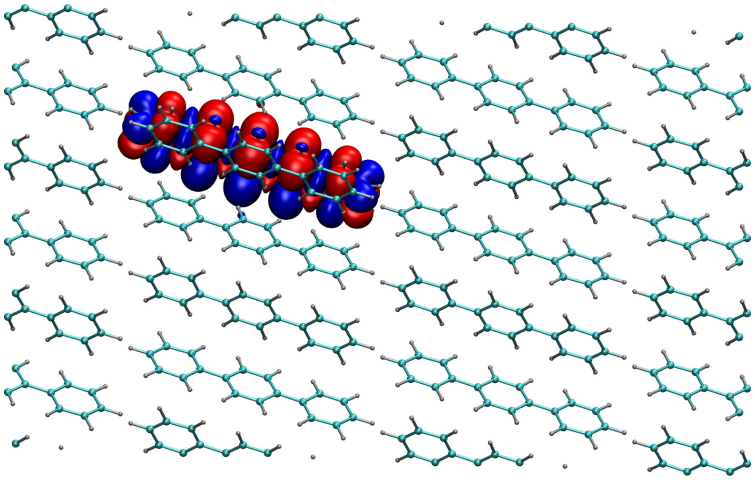
the dopant molecule, while the HOMO-1 and the LUMO+1 state belong to the *p*-terphenyl crystal. The four different systems do have similar Kohn–Sham gaps, ranging from 1.066 eV for c-DNP to 1.261 eV for HNTMP. However, the positioning of the HOMO and LUMO states with respect to the *p*-terphenyl states differs considerably, with the largest difference observed between pentacene and c-DNP. The positioning of the HOMO-LUMO gap with respect to the *p*-terphenyl states has a strong influence over how the exciton delocalises. Since the delocalisation of the excitation must necessarily increase the diagonal part of the excitation energy, a delocalisation is only favourable if it is balanced out by a decrease of the energy associated with the self-consistent potential $\mathbf{V}_{\mathrm{SCF}}^{\{1\}}$. Depending on the strength of the self-consistent field response of the system, a strong coupling to the charge transfer states HOMO→LUMO+1 and HOMO-1→LUMO can become energetically favourable in an ALDA treatment of the system, since these transitions have a relatively low self-consistent field response associated with them. In order to make sure that all excitations have a relatively localised exciton character on the defect molecule, all low energy excited states that have more than 50 % charge-transfer character are ignored in the calculation[44], since their energy is assumed to be unphysically lowered due to the lack of long range exchange interaction in the exchange-correlation kernel.

The excitation energies for the low energy singlet and triplet states that do have less than 50% charge-transfer character and are localised on the dopant molecules are

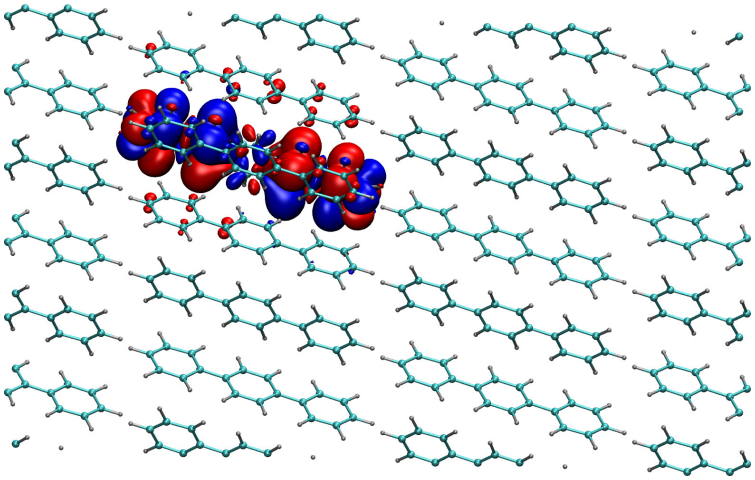
[44]The fraction of charge-transfer character is obtained by determining the dominant Kohn–Sham transitions mixed into the transition density matrix. Since the Kohn–Sham states in the system are either localised on the dopant molecule or on the surrounding *p*-terphenyls, it is possible to identify transitions as charge-transfer states. Furthermore, transitions are compared to the ones obtained if the response density matrix is constrained onto the dopant molecule only.



(a) S_1



(b) T_1



(c) T_2

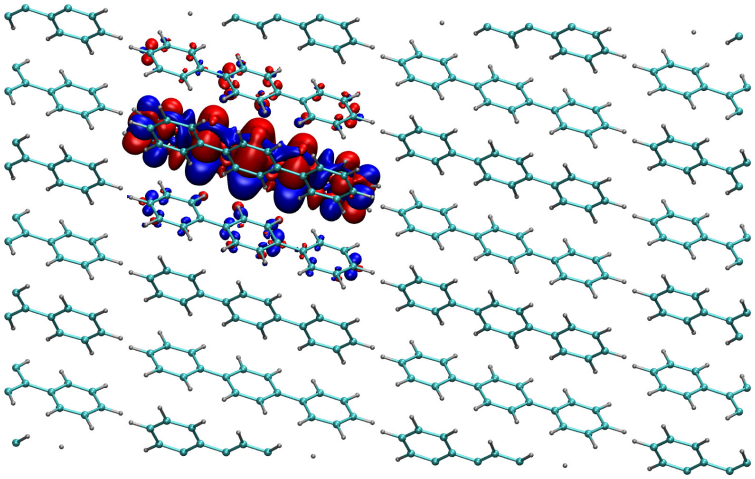
Figure 9.9: Transition densities of the S_1, T_1 and T_2 states of the pentacene dopant molecule.

summarised in tables 9.3, 9.4 and 9.5, where they are compared to their corresponding states for the isolated molecule in vacuum. Furthermore, transition densities of selected excited states for all four dopant molecules in *p*-terphenyl can be found in figures 9.9 to 9.12.

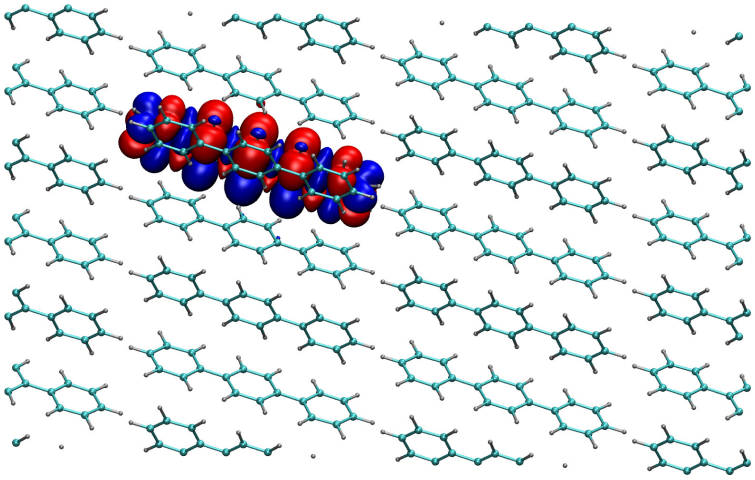
Analysing the results for pentacene (see table 9.3) substituted in *p*-terphenyl, it is found that the T_2 and S_1 states do not precisely line up in the ALDA treatment of the system, with a discrepancy of 0.22 eV in their excitation energy. Furthermore, the T_1 excitation energy is increased by 20 meV in the *p*-terphenyl crystal compared to the vacuum result, while the T_2 state is almost identical. In the S_1 state, an increased delocalisation compared to the triplet states (see figure 9.9) leads to a decrease of the excitation energy by over 50 meV and an increase of the oscillator strength. In general it is thus found that the electrostatic potential of the crystal environment leads to a confinement of excitations and thus an increase of excitation energies, which is counterbalanced by the delocalisation to neighbouring *p*-terphenyl molecules. Depending on the character and energy of the excitation and thus the amount of delocalisation occurring, the net effect can be an increase or a decrease of the excitation energy compared to the isolated molecule in vacuum.

For pentacene, it is helpful to compare the results obtained with experimental results available both in the gas phase and in the *p*-terphenyl crystal. It has long been known that TDDFT significantly underestimates the energy of the S_1 state in vacuum [158, 159], where the experimental value of the S_1 energy in gas phase is known to be 2.23 eV [160]. The experimental results obtained in [154] suggest that this singlet state shifts to an energy of about 2.12 eV when embedding the pentacene in *p*-terphenyl. Thus while the TDDFT calculation presented here underestimates the lowest singlet transition by approximately 0.36 eV in vacuum it predicts the shift of the state when placed in *p*-terphenyl to be in the right direction. While the results obtained here predict a shift of about 53 meV, the experimental results suggest a shift that is twice as large, with about 0.11 eV. This underestimation of the shift is most likely due to the limitation of the spread of the response density matrix to nearest neighbour *p*-terphenyls only, yielding an underestimation in the exciton delocalisation. However, it can be concluded that while the absolute energies of the singlet transition cannot be reliably predicted using TDDFT, the environmental effects of placing the defect molecules into the *p*-terphenyl structure, which are of interest in this computational study, are expected to be predicted with much less of an absolute error.

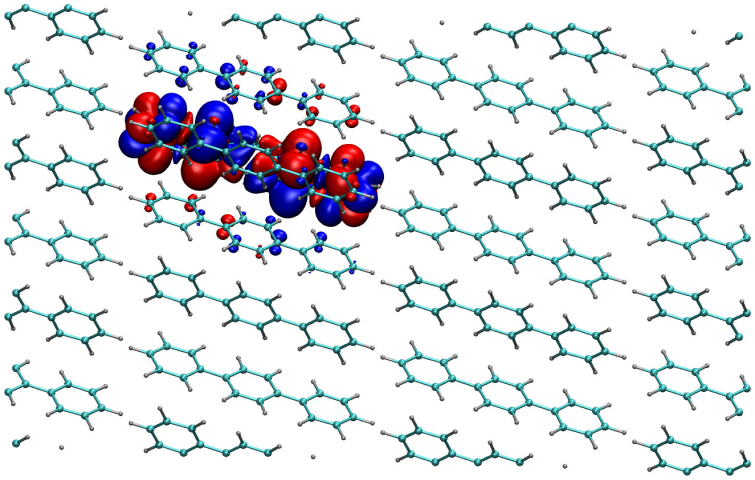
For the o-DNP molecule, it is found that excitation energies in vacuum that are very close to those of the pentacene molecule in vacuum. In *p*-terphenyl, this trend is repeated, in that the T_1 state shows an increase while the S_1 state shows a decrease of the excitation energy. However, the increased delocalisation of the o-DNP S_1 state in *p*-terphenyl leads



(a) S_1

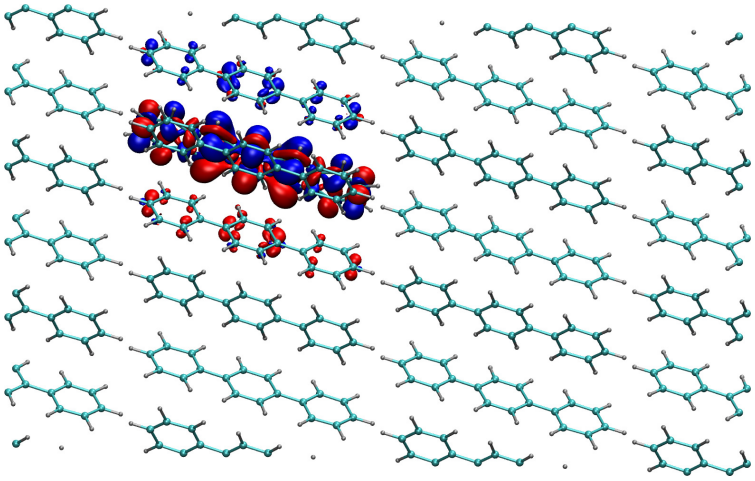


(b) T_1

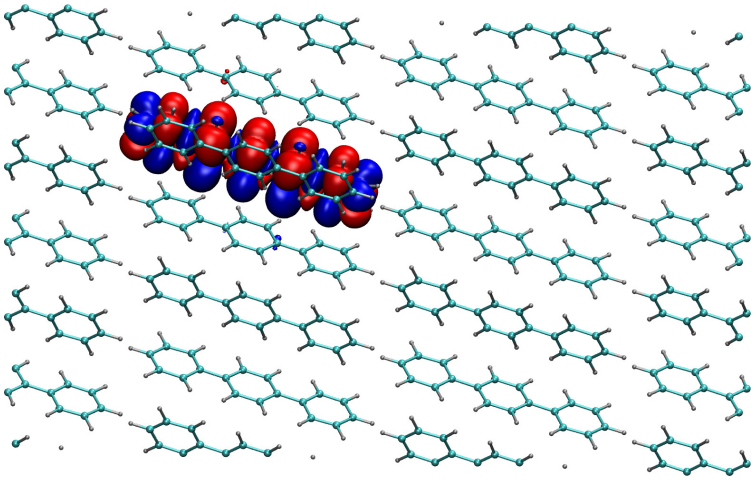


(c) T_2

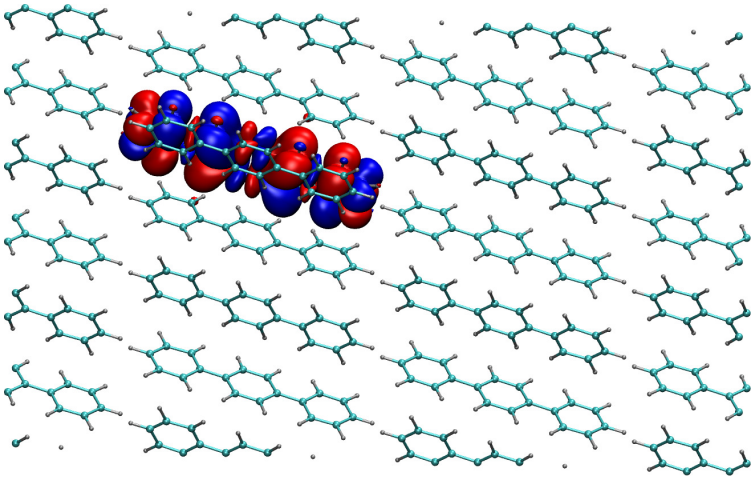
Figure 9.10: Transition densities of the S_1, T_1 and T_2 states of the defect molecule o-DNP.



(a) S_1



(b) T_1



(c) T_3

Figure 9.11: Transition densities of the S_1, T_1 and T_3 states of the dopant molecule c-DNP.

	c-DNP vacuum	c-DNP p-terphenyl
T_1	0.915	0.912
T_2	1.659	1.629
T_3	1.805	1.792
S_1	1.748 (0.0345)	1.749(0.0193)

Table 9.4: Excitation energies for low energy singlet and triplet states of the c-DNP molecule, both in vacuum and *p*-terphenyl. Energies are given in eVs, oscillator strength in brackets.

	HNTMP vacuum	HNTMP *p*-terphenyl
T_1	1.035	1.090
T_2	1.497	1.524
T_3	1.592	1.674
T_4	1.604	1.688
T_5	2.091	2.174
S_1	1.676 (0.0001)	1.703 (0.0007)
S_2	1.684 (0.0002)	1.716 (0.0002)
S_3	1.730 (0.0009)	1.763 (0.0008)
S_4	1.859 (0.0273)	1.964 (0.0320)

Table 9.5: Excitation energies for low energy singlet and triplet states of the HNTMP molecule, both in vacuum and *p*-terphenyl. Energies are given in eVs, oscillator strength in brackets.

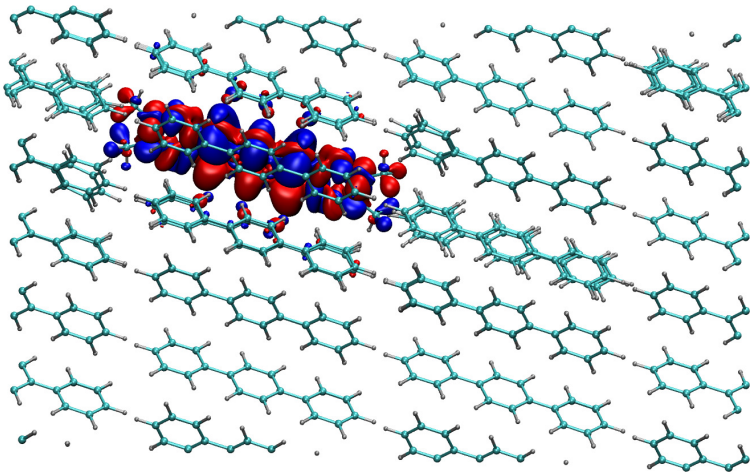
to a drop in energy compared to the vacuum reference state by 0.106 eV, twice as much as in the pentacene dopant. This increased delocalisation for the o-DNP molecule can be explained through changes in the positioning of the HOMO and LUMO as compared to the *p*-terphenyl gap (see table 9.2). The results obtained for o-DNP suggest that it will have very similar characteristics to pentacene when substituted into the *p*-terphenyl crystal. The increased delocalisation of the S_1 state and thus increased energy gap between S_1 and T_2 might worsen the efficiency of the intersystem crossing compared to the pentacene defect, while the introduction of two nitrogen atoms has the potential to influence the triplet splitting of T_1 and the rate of intersystem crossing due to changes in the spin-orbit coupling of the molecule. To the accuracy of the calculations performed here however, it can be concluded that the o-DNP molecule forms a potential candidate to replace pentacene as the photoactive molecule in a room-temperature maser.

In the case of the c-DNP molecule, the close analogy to pentacene changes to a certain extent. First of all, the T_2 state of pentacene and o-DNP has the same character as the T_3 state of c-DNP (see figure 9.9 and 9.11), meaning that the c-DNP molecule gains an additional triplet state between the target state T_1 and the triplet state that couples to S_1 to make the intersystem crossing possible. Furthermore, the excitation energies of c-

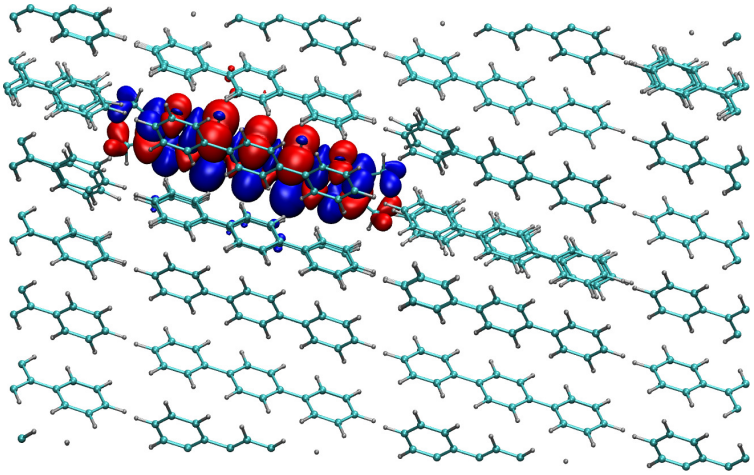
DNP in *p*-terphenyl are in very close agreement with c-DNP in vacuum, with the largest difference being 30 meV for T_2. It is also noted, that the alignment between the S_1 and the T_3 state is much closer than that between S_1 and T_2 in the previous two examples of pentacene and o-DNP, with a difference of only 43 meV. While the standard TDDFT error associated with typical localised excitations on small molecules does not allow us to predict with confidence whether the alignment of singlet and triplet states is indeed more favourable for intersystem crossing in the o-DNP case, it is certainly plausible that the crossing can be achieved in this system. Whether the insertion of an extra triplet state state 0.16 eV below the T_3 state changes the reaction pathway down to the masing state T_1 in a positive or negative way however cannot be predicted from this study.

For the HNTMP molecule, one does not only obtain additional states in the triplet excitations but also the singlet ones. The high oscillator strength singlet state most likely to absorb low energy photons corresponds to S_4 rather than S_1 in the HNTMP molecule. There are three lower energy singlet states, which have a relatively low oscillator strength associated with them. Furthermore, between target state T_1 and the state T_5 most likely to couple with the high oscillator strength state S_4, there are three additional triplet states. Contrary to the previous example molecules, all excitation energies are increased in the *p*-terphenyl crystal compared to the vacuum case, some by more than 0.1 eV. This is most likely due to the more significant distortions experienced by the HNTMP molecule in the *p*-terphenyl crystal compared to other defect molecules, as the forces on the HNTMP molecule are significantly higher. The alignment between the T_5 and S_4 states however is comparable to that of the o-DNP molecule.

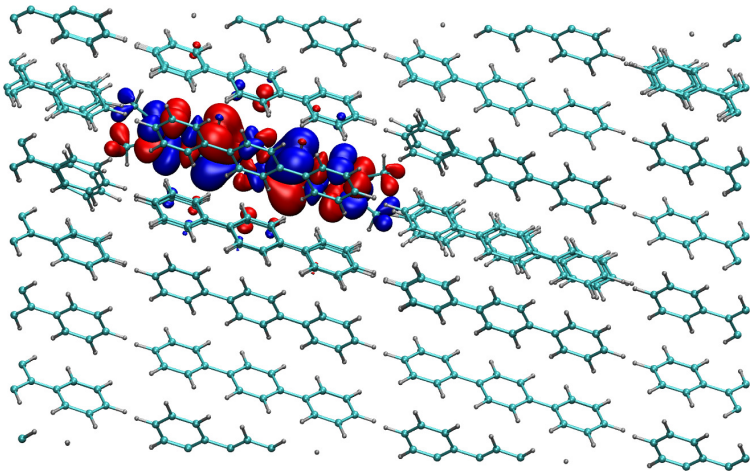
It can be concluded that pentacene in a room-temperature maser can most likely be substituted by o-DNP without significantly altering the efficiency of the pathway to the masing state T_1. The absorption energy and strengths of the two S_1 states for these two molecules in *p*-terphenyl are almost identical. While c-DNP shows the best alignment between the singlet and triplet states for intersystem crossing in the ALDA study performed here, it introduces an additional triplet state between the target state T_1, with unknown consequences for the reaction pathway of the exciton. Furthermore, the oscillator strength of its S_1 state is lower than that of pentacene or o-DNP in *p*-terphenyl by a factor of 3, suggesting that a maser based on c-DNP might be significantly less efficient at absorbing light into the desired singlet state. Of the three pentacene derivatives tested here, HNTMP provides the most doubtful results as a masing molecule. It causes large distortions to the *p*-terphenyl crystal structures and the fact that there are three low lying singlet states below the state coupling strongly to light suggests that there might be significant loss from hot excitons decaying into lower singlet states, if the coupling to singlet states happens at a faster rate to that of intersystem crossing. In any case, the



(a) S_4



(b) T_1



(c) T_5

Figure 9.12: Transition densities of the S_4, T_1 and T_5 states of the defect molecule HNTMP. Note the perturbations in the structure of the nearest neighbour *p*-terphenyl molecules due to the presence of the defect molecule.

additional singlet and triplet states in HNTMP are likely to significantly complicate the reaction pathway from the absorption state S_4 to the masing state T_1, which is likely to have detrimental effects on the efficiency of that reaction.

It is again to be pointed out that the absolute energies obtained in the TDDFT study are not expected to be in very close agreement with experiment, especially for the singlet states. However, as can be seen from the experimental results for pentacene, the absolute shift of excitation energies when placing the molecules from vacuum into the *p*-terphenyl crystal is in much closer agreement with experiment. It is expected that this trend remains true for all other dopant molecules considered here, meaning that the absolute shift energies between the calculation in vacuum and the calculation in *p*-terphenyl can be compared directly to experimental results. It can thus be concluded that the theoretical study performed here can be treated as a first step to prescreening potential dopant molecules for a room-temperature maser.

Chapter 10

Conclusion and future work

In conclusion, an approach has been outlined to compute optical properties of very large systems previously inaccessible with conventional methods. It was argued that while approaches derived from many-body perturbation theory are known to describe excitonic effects in complex systems to a very high degree of accuracy, a good balance between accuracy and computational cost makes time-dependent density-functional theory the only currently available computational method for large systems. This is especially true for systems that already require linear-scaling techniques to converge ground state properties. The linear-scaling DFT code `ONETEP` has been successfully extended to the TDDFT linear-response formalism, and it has been demonstrated that the method scales fully linearly with system size. The method makes use of two different localised orbital representations, one to expand the conduction and one to expand the valence subspace. It has been demonstrated that these two compact sets yield a very efficient representation of the effective transition density matrix of an excitation, allowing for the routine calculation of low energy excitations of systems containing up to 2000 atoms without making use of an explicit truncation of any of the density matrices. It was furthermore demonstrated that the algorithm developed in this work is capable of treating transition metals with great accuracy when extended to the projector augmented-wave formalism.

However, the main insight gained in this work is the way the truncation of the response density matrix can be used to converge targeted excitations in a system. While any truncation of the response density matrix guarantees that the computational cost of converging a single excitation scales linearly with system size, it was outlined in chapter 8 that this does not mean that low energy spectra can be converged in linear scaling effort. However, by truncating the response density matrix in a way to force excitations to localise as described in the subsystem TDDFT approach developed in chapter 8, a fully linear scaling calculation of low energy spectra composed of coupled localised excitons is possible, opening up a wide range of applications from doped molecular crystals to large scale photoactive biological systems.

In chapter 9, explicit calculations on real systems of interest were outlined both in the area of biophysics and that of molecular crystals, and the power of the subsystem TDDFT approach was demonstrated. It was shown that in order to converge localised excitons in the Fenna-Matthews-Olson complex, it is crucial to treat a large part of the protein environment of a chromophore fully quantum mechanically, making large scale calculations such as the ones enabled by the linear-scaling TDDFT approach presented here necessary. Furthermore, in the second physically motivated system treated in this work, the subsystem TDDFT approach was used to study excitons localised on defect molecules within a *p*-terphenyl crystal. Together with structural information gained from linear-scaling DFT, three different defect molecules were analysed and compared to pentacene in their suitability for being used in a room-temperature zero-field maser. It was possible to identify two of the molecules as likely candidates with similar exciton characteristics as pentacene when placed into a *p*-terphenyl crystal, while one molecule could be ruled out as ineffective. It is to be pointed out that most of the large scale systems treated in this work have sizes of the order of 2000 atoms, which is at the lower end of the capabilities of linear-scaling DFT. While systems of this moderate size are already beyond the capabilities of most conventional cubic-scaling TDDFT methods, a real test of the linear-scaling TDDFT approach developed here will come in the form of larger systems. In chapter 9 one such application was already proposed, where the low energy optical absorption spectrum of the entire Fenna-Matthews-Olson complex could be computed without having to resort to any semi-classical treatment of parts of the system. The techniques developed in this work open up a wide range of potential applications unaccessible by previous methods and have shown the potential to bridge the gap between computational simulations and experiment.

While the findings outlined in this work lay the foundations for some interesting large scale applications of TDDFT, there are a number of desirable further developments that will be addressed in future work. The extension of the algorithm to hybrid functionals via techniques to evaluate Hartree-Fock exchange in linear scaling effort is crucial in describing long-range charge transfer states and excitonic effects in solids. Furthermore, an extension of the two-step minimisation approach of the ground state DFT method to TDDFT is desirable, as it will allow for the optimisation of generalised response orbitals to represent the response density matrices of some low lying excitation, rather than relying on an explicit conduction optimisation for a chosen number of unoccupied states. From a theoretical point of view, the explicit minimisation of the TDDFT energy functional with respect to some generalised response NGWFs would allow for the rigorous definition of the response and valence NGWFs as the optimum representation for a given set of excitations and a given localisation constraint. Finally, it is desirable to extend the formalism to allow

for the linear-scaling evaluation of excited state forces, which will open up the possibility of excited state geometry optimisations of very large systems, with potential applications to the calculation of photoemission spectra and transition states of large photoactive biomolecules.

Bibliography

[1] Dirac, P. A. M. Quantum Mechanics of Many-Electron Systems. *Proc. R. Soc. B* **123**, 714-733 (1929).

[2] Wentzel, G. Eine Verallgemeinerung der Quantenbedingungen für die Zwecke der Wellenmechanik. *Z. Phys.* **38**, 518–529 (1926).

[3] Kramers, H. A. Wellenmechanick und halbzählige Quantisierung. *Z. Phys.* **39**, 828–840 (1926).

[4] Brillouin, L. La méchanique ondulatoire de Schrödinger: une méthode générale de resolution par approximations successives. *C. R. Acad. Sci.* **183**, 24–26 (1926).

[5] Jeffreys, H. On certain approximate solutions of linear differential equations of second order. *Proc. London Math. Soc.* **23**, 428–436 (1924).

[6] Gamov, G. & Teller, E. The Rate of Selective Thermonuclear Fusion. *Phys. Rev.* **53**, 608–609 (1938).

[7] Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864-B871 (1964).

[8] Heiss, M., Fontana, Y., Wüst, G., Magen, C., O'Reagan, D. D., Luo, J. W., Ketterer, B., Conesa-Boj, S., Kuhlmann, A. V., Houel, J., Russo-Averchi, E., Morante, J. R., Cantoni, M., Marzari, N., Arbiol, J., Zunger, A., Warburton, R. J. & Fontcuberta i Morral, A. Self-assembled quantum dots in a nanowire system for quantum photonics. *Nat. Mater.* **12**, 439-444 (2013).

[9] Cole, D., O'Reagan, D. D. & Payne, M. C. Ligand Discrimination in Myglobin from Linear-Scaling DFT+U. *J. Phys. Chem. Lett.* **3**, 1448 (2012).

[10] Miyazaki, T., Bowler, D. R., Gillan, M. J. & Ohno, T. The Energetics of Hut-Cluster Self-Assembly in Ge/Si(001) from Linear-Scaling DFT Calculations. *J. Phys. Soc. Jpn.* **77**, 123706 (2008).

[11] Ordejón, P., Linear Scaling ab initio Calculations in Nanoscale Materials with SIESTA. *Phys. Status Solidi b* **217**, 335-356 (2000).

[12] McQuarrie, D. A. *Quantum Chemistry* (University Science Books, 2008).

[13] Dirac, P. A Theory of Electrons and Protons. *Proc. R. Soc. A* **126**, 360 (1927).

[14] Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **84**, 457 (1927).

[15] Ziman, J. M. *Principles of the Theory of Solids* (Cambridge University Press, 1972).

[16] Slater, J. & Verma, H. C. The Theory of Complex Spectra. *Phys. Rev.* **32**, 1293-1295 (1929).

[17] Møller, C. & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **46**, 618 (1934).

[18] Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **45**, 4256 (1966).

[19] Szalay, P. G., Müller, T., Gidofalvi, G., Lischka, H. & Shepard, R. Multiconfiguration Self-Consistent Field and Multireference Configuration Interaction Methods and Applications. *Chem. Rev.* **112**, 108-181 (2012).

[20] Levy, M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and the solution to the v-representability problem. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 6062 (1979).

[21] Gilbert, T. L. Hohenberg-Kohn theorem for nonlocal external potentials. *Phys. Rev. B* **12**, 2111-2120 (1975).

[22] Harriman, J. E. Orthonormal orbitals for the representation of an arbitrary density. *Phys. Rev. A* **24**, 680-682 (1981).

[23] Thomas, L. H. The calculation of atomic fields. *Proc. Cambridge Philos. Soc.* **23**, 542 (1927).

[24] Fermi, E. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. *Z. Phys.* **48**, 73 (1928).

[25] von Weizsäcker, C. F. Zur Theorie der Kernmassen. *Z. Phys.* **96**, 431 (1935).

[26] Chai, J.-D. & Weeks, J. D. Orbital-Free Density Functional Theory: Kinetic Potentials and Ab Initio Local Pseudopotentials. *Phys. Rev. B* **75**, 205122 (2007).

[27] Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, 1133 (1965).

[28] Kresse, G. & Furthmüller, J. Efficient schemes for *ab-initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).

[29] Smarglassl, E. & Madden, P. A. Orbital-free kinetic energy functionals for first-principles molecular dynamics. *Phys. Rev. B* **49**, 5220 (1994).

[30] Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms. *Physica* **1**, 104-113 (1934).

[31] Perdew, J. P., Parr, R. G., Levy, M. & Balduz Jr., J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **49**, 1691 (1982).

[32] Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critcal analysis. *Can. J. Phys.* **58**, 1200 (1980).

[33] Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048 (1981).

[34] Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **45**, 13244 (1992).

[35] Ceperley, C. M. & Adler, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **45**, 566 (1980).

[36] Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **140**, 18A301 (2014).

[37] Becke, A. D. Density functional calculations of molecular bond energies. *J. Chem. Phys.* **84**, 4524 (1986).

[38] Perdew, J. P. & Wang, Y. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B* **33**, 8800 (1986).

[39] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865 (1996).

[40] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J Chem. Phys.* **98**, 5648 (1993).

[41] Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982 (1996).

[42] Dion, M., Rydberg, H., Schröder, E., Langreth, D. C. & Lundquist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **92**, 246401 (2004).

[43] Bloch, F. Über die Quantenmechanik der Elektronen in Kristallgittern. *Z. Phys.* **52**, 555-600 (1929).

[44] Marzari, N. & Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **56**, 12847 (1997).

[45] Kohn, W. Analytic Properties of Bloch Waves and Wannier Functions. *Phys. Rev.* **115**, 809 (1959).

[46] des Cloizeaux, J. Analytical Properties of n-Dimensional Energy Bands and Wannier Functions. *Phys. Rev.* **135**, A698 (1964).

[47] Nenciu, G. Existence of Exponentially Localised Wannier Functions. *Commun. Math. Phys.* **91**, 81 (1983).

[48] Brouder, C., Panati, G., Calandra, M., Mourougane, C. & Marzari, N. Exponential Localization of Wannier Functions in Insulators. *Phys. Rev. Lett.* **98**, 046402 (2007).

[49] Cooley, J. W. & Tukey, J. W. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.* **19**, 297-301 (1965).

[50] Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **369**, 253-287 (1921).

[51] Arbenz, P., Hetmaniuk, U. L., Lehoucq, R. B. & Tuminaro, R. S. A comparison of eigensolvers for large-scale 3D modal analysis using AMG-preconditioned iterative methods. *Int. J. Numer. Meth. Eng.* **64**, 204-236 (2005).

[52] Davidson, E. R. The iterative calculation of a few lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.* **17**, 817-953 (1975).

[53] Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64**, 1045-1097 (1992).

[54] Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **43**, 1993 (1991).

[55] Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).

[56] Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **71**, 1085 (1999).

[57] Bowler, D. R. & Miyazaki, T. $\mathcal{O}(N)$ methods in electronic structure calculations. *Rep. Prog. Phys.* **75**, 036503 (2012).

[58] McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **32**, 335 (1960).

[59] Junquera, J., Paz, O., Sánchez-Portal & Artacho, E. Numerical atomic orbitals for linear-scaling calculations. *Phys. Rev. B* **64**, 235111 (2001).

[60] Anglada, E., Soler, J. M., Junquera, J. & Artacho, E. Systematic generation of finite-range atomic basis sets for linear-scaling calculations. *Phys. Rev. B* **66**, 205101 (2002).

[61] Artacho, E. & Miláns del Bosch, L. Nonorthogonal basis sets in quantum mechanics: Representations and second quantization. *Phys. Rev. A* **43**, 5770 (1991).

[62] Head-Gordon, M., Maslen, P. E. & White, C. A. A tensor formulation of many-electron theory in a nonorthogonal single-particle basis. *J. Chem. Phys.* **108**, 616 (1998).

[63] Li, X.-P., Nunes, R. W. & Vanderbilt, D. Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B* **47**, 10891 (1993).

[64] Kohn, W. Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms. *Phys. Rev. Lett.* **76**, 3168 (1996).

[65] Haynes, P. D. & Payne, M. C. Failure of density-matrix minimization methods for linear-scaling density-functional theory using the Kohn penalty-functional. *Solid State Commun.* **108**, 737-741 (1998).

[66] Haynes, P. D. & Payne, M. C. Corrected penalty-functional method for linear-scaling calculations within density-functional theory. *Phys. Rev. B* **59**, 12173 (1999).

[67] Ismail-Beigi, S. & Arias, T. A. Locality of the Density Matrix in Metals, Semiconductors, and Insulators. *Phys. Rev. Lett.* **82**, 2127 (1999).

[68] Franco de Carvalho, F., Curchod, B. F. E., Penfold, T. J. & Tavernelli, I. Derivation of spin-orbit couplings in collinear linear-response TDDFT: A rigorous formulation. *J. Chem. Phys.* **140**, 144103 (2014).

[69] Runge, E. & Gross, E. K. U. Density Functional Theory for Time-Dependent systems. *Phys. Rev. Lett.* **52**, 997 (1984).

[70] Ghosh, S. K. & Dhara, A. K. Density-functional theory of many-electron systems subjected to time-dependent electric and magnetic fields. *Phys. Rev. A* **38**, 1149 (1988).

[71] Vignale, G. Mapping from current densities to vector potentials in time-dependent current density functional theory. *Phys. Rev. B* **70**, 201102(R) (2004).

[72] Kohl, H. & Dreizler, R. M. Time-Dependent Density-Functional Theory: Conceptual and Practical Aspects. *Phys. Rev. Lett.* **56**, 1993 (1986).

[73] van Leeuwen, R. Causality and Symmetry in Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **80**, 1280 (1998).

[74] Marques, M. A. L. & Gross, E. K. U. Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.* **55**, 427 (2004).

[75] Yabana, K. & Bertsch, G. F. Time-dependent local-density approximation in real time. *Phys. Rev. B* **54**, 4484 (1996).

[76] Castro, A., Marques, M. A. L. & Rubio, A. Propagators for the time-dependent Kohn-Sham equations. *J. Chem. Phys.* **121**, 3425 (2004).

[77] Onida, G., Reining, L. & Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **74**, 601 (2002).

[78] Petersilka, M., Grossmann, U. J. & Gross, E. K. U. Excitation Energies from Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **76**, 1212 (1996).

[79] Casida, M. E., in *Recent Advances in Density Functional Methods Vol. 1*, edited by Chong, D. P. (World Scientific, Singapore, 1995)

[80] Hübener, H. & Giustino, F. Linear optical response of finite systems using multishift linear system solvers. *J. Chem. Phys.* **141**, 044117 (2014).

[81] Berkowitz, J. *Photoabsorption, Photoionization, and Photoelectron Spectroscopy* (Academic Press: New York, 1979)

[82] Fetter, A. L. & Walecka, J. D. *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971)

[83] Hirata, S. & Head-Gordon, M. Time-dependent density functional theory within the Tamm-Dancoff approximation. *Chem. Phys. Lett.* **314**, 291-299 (1999).

[84] Rohlfing, M. & Louie, S. G. Electron-hole excitations and optical spectra from first principles. *Phys. Rev. B* **62**, 4927 (2000).

[85] Casida, M. E. Time-dependent density-functional theory for molecules and molecular solids. *J. Mol. Struc.-THEOCHEM* **914**, 3-18 (2009).

[86] Peach, M. J. G., Benfield, P., Helgaker, T. & Tozer, D. J. Excitation energies in density functional theory: An evaluation and a diagnostic test. *J. Chem. Phys.* **128**, 044118 (2008).

[87] Sadovskii, M. V., *Diagrammatics: Lectures on Selected Problems in Condensed Matter Theory*, (World Scientific Publishing Co Pte Ltd, Singapore, 2006)

[88] Lifshitz, E. M. & Pitaevskii, L. P., *Statistical Physics: Theory of the Condensed State (Course of Theoretical Physics Vol. 9)*, (Pergamon Press, Oxford, 1980)

[89] Hedin, L. New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem. *Phys. Rev.* **139**, A796 (1965).

[90] Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B* **34**, 5390 (1986).

[91] Brar, V. W., Wickenburg, S., Panlasigui, M., Park, C.-H., Wehling, T. O., Zhang, Y., Decker, R., Girit, Ç., Balatsky, A. V., Louie, S. G., Zettl, A. & Crommie, M. F. Observation of Carrier-Density-Dependent Many-Body Effects in Graphene via Tunneling Spectroscopy. *Phys. Rev. Lett.* **104**, 036805 (2010).

[92] Reining, L., Olevano, V., Rubio, A. & Onida, G. Excitonic Effects in Solids Described by Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **88**, 066404 (2002).

[93] Foerster, D., Koval, P. & Sánches-Portal, D. An $O(N^3)$ implementation of Hedin's GW approximation for molecules. *J. Chem. Phys.* **135**, 074105 (2011).

[94] Umari, P., Stenuit, G. & Baroni, S. GW quasiparticle spectra from occupied states only. *Phys. Rev. B* **81**, 115104(2010).

[95] Giustino, F., Cohen, M. L. & Louie, S. G. GW method with the self-consistent Sternheimer equation. *Phys. Rev. B* **81**, 115105 (2010).

[96] Lambert, H. & Giustino, F. Ab initio Sternheimer-GW method for quasiparticle calculations using plane waves. *Phys. Rev. B* **88**, 075117 (2013).

[97] Yam, C., Yokojima, S. & Chen, G. Linear-scaling time-dependent density-functional theory *Phys. Rev. B* **68**, 153105 (2003).

[98] Tretiak, S., Isborn, C. M., Niklasson, A. M. N. & Challacombe, M. Representation independent algorithms for molecular response calculations in time-dependent self-consistent field theories. *J. Chem. Phys.* **130**, 054111 (2009).

[99] Zuehlsdorff, T. J., Hine, N. D. M., Spencer, J. S., Harrison, N. M., Riley, D. J. & Haynes, P. D. Linear-scaling time-dependent density-functional theory in the linear response formalism. *J. Chem. Phys.* **139**, 064104 (2013).

[100] Challacombe, M. Linear Scaling Solution of the Time-Dependent Self-Consistent-Field Equations. *Computation* **2**, 1-11 (2014).

[101] Haynes, P. D., Skylaris, C.-K., Mostofi, A. A. & Payne, M. C. ONETEP: linear-scaling density functional theory with plane waves. *Psi-k Newsletter* **72**, 78-91 (2005).

[102] Haynes, P. D., Skylaris, C.-K., Mostofi, A. A. & Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **122**, 084119 (2005).

[103] Mostofi, A. A., Skylaris, C.-K., Haynes, P. D. & Payne, M. C. Total-energy calculations on a real space grid with localized functions and a plane-wave basis. *Comput. Phys. Commun.* **147**, 788 (2002).

[104] Skylaris, C.-K., Mostofi, A. A., Haynes, P. D., Diéguez, O. & Payne, M. C. Nonorthogonal generalized Wannier function pseudopotential plane-wave method. *Phys. Rev. B* **66**, 035119 (2002).

[105] Skylaris, C.-K., Mostofi, A. A., Haynes, P. D., Pickard, C. J. & Payne, M. C. Accurate kinetic energy evaluation in electronic structre calculations with localized functions on real space grids. *Comput. Phys. Commun.* **140**, 315-322 (2001).

[106] Hine, N. D. M., Haynes, P. D., Mostofi, A. A., Skylaris, C.-K. & Payne, M. C. Linear-scaling density-functional theory with tens of thousands of atoms: Expanding the scope and scale of calculations with ONETEP. *Comput. Phys. Commun.* **180**, 1041-1053 (2009).

[107] Hine, N. D. M., Robinson, M., Haynes, P. D. Skylaris, C.-K., Payne , M. C. & Mostofi, A. A. Accurate ionic forces and geometry optimizations in linear-scaling density-functional theory with local orbitals. *Phys. Rev. B* **83**, 195102 (2011).

[108] Ruiz-Serrano, A., Hine, N. D. M. & Skylaris, C.-K. Pulay forces from localized orbitals optimized *in situ* using a psinc basis set. *J. Chem. Phys.* **136**, 234101 (2012).

[109] Ratcliff, L. E., Hine, N. D. M. & Haynes, P. D. Calculating optical absorption spectra for large systems using linear-scaling density functional theory. *Phys. Rev. B* **84**, 165131 (2011).

[110] Kresse, G. & Joubert, D. From ultrasoft pesudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).

[111] Torrent, M., Joliet, F., Bottin, F. *et al.* Implementation of the projector augmented-wave method in the ABINIT code: Applications to the study of iron under pressure. *Comp. Mater. Sci.* **42**, 337-351 (2008).

[112] Lever, G., Cole, D. J., Hine, N. D. M., Haynes, P. D. & Payne, M. C. Electrostatic considerations affecting the calculated HOMO-LUMO gap in protein molecules. *J. Phys.: Condens. Matter* **25**, 152101 (2013).

[113] Cramer, C. J. & Truthlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **99**, 2161-2200 (1999).

[114] Dziedzic, J., Helal, H. H., Skylaris, C.-K., Mostofi, A. A. & Payne, M. C. Minimal parameter implicit solvent model for ab initio electronic-structure calculations. *Europhys. Lett.* **95**, 43001 (2011).

[115] Cococcioni, M., Mauri, F., Ceder, G. & Marzari, N. Electronic-Enthalpy Functional for Finite Sytems Under Pressure. *Phys. Rev. Lett.* **94**, 145501 (2005).

[116] Corsini, N. R. C., Greco, A., Hine, N. D. M., Molteni, C. & Haynes, P. D. Simulations of nanocrystals under pressure: Combining electronic enthalpy and linear-scaling density-functional theory. *J. Chem. Phys.* **139**, 084117 (2013).

[117] Hutter, J. Excited state nuclear forces from the Tamm-Dancoff approximation to time-dependent density functional theory within the plane wave basis set framework. *J. Chem. Phys.* **118**, 3928 (2003).

[118] Tsiper, E. V. A classical mechanics technique for quantum linear response. *J. Phys. B* **34**, L401 (2001).

[119] Furche, F. On the density matrix based approach to time-dependent density functional response theory. *J. Chem. Phys.* **114**, 5982 (2001).

[120] Gan, C. K., Haynes, P. D. & Payne, M. C. Preconditioned conjugate gradient method for the sparse generalized eigenvalue problem in electronic structure calculations. *Comput. Phys. Commun.* **134**, 33-40 (2001).

[121] Souza, I., Marzari, N. & Vanderbilt, D. Maximally localized Wannier functions for entangled energy bands. *Phys. Rev. B* **65**, 035109 (2001).

[122] Chen, G. & Mukamel, S. Nonlinear Polarizabilities of Donor-Acceptor Substituted Conjugated Polyenes *J. Phys. Chem.* **100**, 11080-11085 (1996).

[123] Hübener, H. & Giustino, F. Time-dependent density functional theory using atomic orbitals and the self-consistent Sternheimer equation. *Phys. Rev. B* **89**, 085129 (2014).

[124] Rocca, D., Gebauer, R., Saad, Y. & Baroni, S. Turbo charging time-dependent density-functional theory with Lanczos chains. *J. Chem. Phys.* **128**, 154105 (2008).

[125] Ekström, U., Visscher, L., Bast, R., Thorvaldsen, A. J. & Ruud, K. Arbitrary-Order Density Functional Response Theory form Automatic Differentiation. *J. Chem. Theory Comput.* **6**, 1971-1980 (2010).

[126] Schwegler, E. & Challacombe, M. Linear scaling computation of the Hartree-Fock exchange matrix. *J. Chem. Phys.* **105**, 2726 (1996).

[127] Dziedzic, J., Hill, Q. & Skylaris, C.-K. Linear-scaling calculation of Hartree-Fock exchange energy with non-orthogonal generalised Wannier functions. *J. Chem. Phys.* **139**, 214103 (2013).

[128] Valiev, M., Bylaska, E. J., Govind, N., Kowalski, K., Straatsma, T. P., van Dam, H. J. J., Wang, D., Nieplocha, J., Apra, E., Windus, T. L. & de Jong, W. A. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **181**, 1477 (2010).

[129] Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007 (1989).

[130] Bauernschmitt, R., Ahlrichs, R., Heinrich, F. H. & Kappes, M. M. Experiment versus Time Dependent Density Functional Theory Prediction of Fullerene Electronic Absorption. *J. Am. Chem. Soc.* **120**, 5052-5059 (1998).

[131] Du, H., Fuh, R. C. A., Li, J., Corkan, L. A. & Lindsey, J. S. PhotochemCAD: A computer-aided design and research tool in photochemistry. *J. Photochem. Photobiol.* **68**, 141 (1998).

[132] Sundholm, D. Density functional theory calculations of the visible spectrum of chlorophyll *a*. *Chem. Phys. Lett.* **302**, 480-484 (1999).

[133] Wang, Y.-L. & Wu, G.-S. Improving the TDDFT calculation of low-lying excited states for polycyclic aromatic hydrocarbons using the Tamm-Dancoff approximation. *Int. J. Quant. Chem.* **108**, 430-439 (2008).

[134] Scalmani, G., Frisch, M. J., Mennucci, B., Tomasi, J., Cammi, R, & Barone, V. Geometries and properties of excited states in the gas phase and in solution: Theory and application of a time-dependent density functional theory polarizable continuum model. *J. Chem. Phys.* **124**, 094107 (2006).

[135] Tronrud, D. E., Wen, J., Gay, L. & Blankenship, R. E. The Structural Basis for the Difference in Absorbance Spectra for the FMO Antenna Protein from Various Green Sulfur Bacteria. *Photosynth. Res.* **100**, 79-87 (2009).

[136] Avraam, P. W., Hine, N. D. M., Tagney, P. & Haynes, P. D. Factors influencing the dispersion of charge in polar nanorods. *Phys. Rev. B* **83**, 241402(R) (2011).

[137] Avraam, P. W., Hine, N. D. M., Tagney, P. & Haynes, P. D. Fermi-level pinning can determine polarity in semiconductor nanorods. *Phys. Rev. B* **85**, 115404 (2012).

[138] Walter, M., Häkkinen, H., Lehtovaara, L., Puska, M., Enkovaara, J., Rostgaard, C. & Mortensen J. J. Time-dependent density-functional theory in the projector augmented-wave method. *J. Chem. Phys.* **128**, 244101 (2008).

[139] Joilet, F., Torrent, M. & Holzwarth, N. Generation of Projector Augmented-Wave atomic data: A 71 element validated table in the XML format. *Comput. Phys. Commun.* **185**, 1246 (2014).

[140] Choi, C. L., Koski, K. J., Sivasankar, S. & Alivisatos, A. P. Strain-Dependent Photoluminescence Behaviour of CdSe/CdS Nanocrystals with Spherical, Linear, and Branched Topologies. *Nano Lett.* **9**, 3544 (2009).

[141] Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **14**, 33-38 (1996).

[142] Polizzi, E. Density-Matrix-Based Algorithms for Solving Eigenvalue Problems. *Phys. Rev. B* **79**, 115112 (2009).

[143] Scholes, G. D., Fleming, G. R., Olaya-Castro, A. & van Grondelle, R. Lessons from Nature about Solar Light Harvesting. *Nat. Chem.* **3**, 763 (2011).

[144] Lambert, N., Chen, Y.-N., Cheng, Y.-C., Li, C.-M., Chen, Y.-G. & Nori, F. Quantum Biology. *Nat. Phys.* **9**, 10-18 (2012).

[145] Adolphs, J., Müh, F., Madjet, M. E. & Renger, T. Pigment Transiton Energies in the FMO Protein. *J. Am. Chem. Soc.* **132**, 3331 (2010).

[146] Adolphs, J., Müh, F., Madjet, M. E., Schmidt am Busch, M. & Renger, T. Structure-Based Calculations of Optical Spectra of Photosystem I suggests an Asymmetric Light-Harvesting Process. *J. Am. Chem. Soc.* **132**, 3331 (2010).

[147] Neugebauer, J. Subsystem-Based Theoretical Spectroscopy of Biomolecules and Biomolecular Assemblies. *Chem. Phys. Chem.* **10**, 3148 (2009).

[148] Neugebauer, J. & Baerends, E. J. Exploring the Ability of Frozen-Density Embedding to Model Induced Circular Dichorism. *J. Phys. Chem. A* **110**, 8786 (2006).

[149] Neugebauer, J. Coupling between electronic transitions in a subsystem formulation of time-dependent density functional theory. *J. Chem. Phys.* **126**, 134116 (2007).

[150] König, C., Schlüter, N. & Neugebauer, J. Direct determination of exciton couplings from subsystem time-dependent density-functional theory within the Tamm-Dancoff approximation. *J. Chem. Phys.* **138**, 034104 (2013).

[151] Fenna, R. E. & Matthews, B. W. Chlorophyll arrangement in a bacteriochlorophyll protein from *Chlorobium limicola*. *Nature* **258**, 573 (1975).

[152] Cole, D. J., Chin, A. W., Hine, N. D. M., Haynes, P. D. & Payne, M. C. Toward Ab Initio Optical Spectroscopy of the Fenna-Matthews-Olson Complex. *J. Phys. Chem. Lett.* **4**, 4206 (2013).

[153] Isborn, C. M., Götz, A. W., Clark, M. A., Walker, R. C. & Martinéz, T. J. Electronic Absorption Spectra from MM and *ab initio* QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **8**, 5092-5106 (2012).

[154] Oxborrow, M., Breeze, J. D. & Alford, N. M. Room-temperature solid-state maser. *Nature* **488**, 353-356 (2012).

[155] Rietveld, H. M., Malsen, E. N. & Clews, C. J. B. An X-ray and Neutron Diffraction Refinement of the Structure of p-Terphenyl. *Acta Cryst.* **B26**, 693 (1970).

[156] Pope, M. & Swenberg, C. E. *Electronic Processes in Organic Crystals and Polymers, 2nd ed.* (Oxford University Press, New York, 1999).

[157] Cudazzo, P., Gatti, M. & Rubio, A. Excitons in molecular crystals from first-principles many-body perturbation theory: Picene versus pentacene. *Phys. Rev. B* **86**, 195307 (2012).

[158] Grimme, S. & Parac, M. Substantial Errors from Time-Dependent Density Functional Theory for the Calculation of Excited States of Large π Systems. *ChemPhysChem* **4**, 292-295 (2003).

[159] Kadantsev, E. S., Scott, M. J. & Rubio, A. Electronic structure and excitations in oligoacenes from *ab initio* calculations. *J. Chem. Phys.* **124**, 134901 (2006).

[160] Biermann, D. & Schmidt, W. Diels-Alder Reactivity of Polycyclic Aromatic Hydrocarbons. 1. Acenes and Benzologs. *J. Am. Chem. Soc.* **102**, 3163 (1980)